



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 142881

TO: Jeffrey Parkin
Location: REM 3d39 / 3c18
Art Unit: 1648
Tuesday, January 25, 2005

Case Serial Number: 09/319156

From: Deirdre Arnold
Location: Biotech-Chem Library
REM 1A64
Phone: 571-272-2532

Deirdre.Arnold@uspto.gov

Search Notes

Please contact me if you need to search the pending files.

Please contact me if you encounter any problems with the file or the disk.

Please feel free to contact me if you have any questions or would like to amend the search.

Thank you for using STIC services.

Regards,
Deirdre Arnold

This Page Blank (uspio)

STIC-Biotech/ChemLib

~~142354~~ 142881

From: Parkin, Jeffrey
Sent: Tuesday, January 11, 2005 8:50 PM
To: STIC-Biotech/ChemLib
Subject: SEQUENCE SEARCH: U.S. Serial No. 09/319,156

Please search **SEQ ID NOS.: 6, 9, 10, and 12** from the aforementioned application (U.S. Serial No. **09/319,156**) v. all relevant databases.

Please provide the first **30 SUMMARIES**.

Place results on both paper and disk.

Thanks!

JSP
AU 1648
REM-3D39
2-0908

http://expoweb1:8001/cgi-bin/expo/GenInfo/snquery.pl?APPL_ID=09319156.

STAFF USE ONLY

Searcher: Arnold
Searcher Phone: 2-2532
Date Searcher Picked up: 1/19/05
Date Completed: 1/25/05
Searcher Prep/Rev. Time: _____
Online Time: _____

Type of Search

NA Sequence: # 3
AA Sequence: # 1
Structure: # _____
Bibliographic: _____
Litigation: _____
Patent Family: _____
Other: _____

Vendors and cost where applicable

STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other(Specify): _____

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2005, 10:09:14 ; Search time 2663.68 Seconds
(without alignments)
11273.477 Million cell updates/sec

Title: US-09-319-156B-6
Perfect score: 635
Sequence: 1 cccgtatctttaacctctc.....tgagaaaaa.....

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- GenEmbl.*
- 1: gb_ba.*
 - 2: gb_hgt.*
 - 3: gb_in.*
 - 4: gb_om.*
 - 5: gb_ov.*
 - 6: gb_pat.*
 - 7: gb_ph.*
 - 8: gb_pl.*
 - 9: gb_pr.*
 - 10: gb_ro.*
 - 11: gb_sts.*
 - 12: gb_sy.*
 - 13: gb_un.*
 - 14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	635	100.0	635	6	AR344385 Sequence
2	635	100.0	635	6	AX001024 Sequence
3	635	100.0	635	6	BD136195 Retrovira
4	635	100.0	2004	14	AF127229 Multiple
5	631.8	99.5	167694	9	AC113137 Homo sapi
6	631.8	99.5	176095	2	AC019346 Homo sapi
7	631.8	99.5	210336	2	AC022171 Homo sapi
8	580.2	91.4	148724	2	AC010778 Homo sapi
9	579.2	91.2	8339	2	AL162912 Human DNA
10	571	89.9	136901	9	AC073626 Homo sapi
11	566.4	89.2	176188	9	AC068898 Homo sapi
12	564.6	88.9	89728	9	AL583805 Human DNA
13	560.2	88.2	112405	9	AC072023 Homo sapi
14	559.8	88.2	179114	9	BS000045 Pan trogl
15	556.6	87.7	174019	9	AP001538 Homo sapi
16	556.6	87.7	340000	9	AP001674 Homo sapi
17	551.8	86.9	149755	2	AP001545 Homo sapi
18	551.8	86.9	152980	2	AC090313 Homo sapi
19	551.8	86.9	163803	9	AC093531 Homo sapi

C	20	551.8	86.9	176425	9	AC107075 Homo sapi
C	21	551.4	86.8	128468	9	AL139090 Human DNA
C	22	551.4	86.8	140756	9	AL139038 Human DNA
C	23	551.4	86.8	167366	2	AC021774 Homo sapi
C	24	551.4	86.8	183499	2	AL607153 Homo sapi
C	25	550.2	86.6	164211	2	AP002790 Homo sapi
C	26	550.2	86.6	167843	9	AC064801 Homo sapi
C	27	549	86.5	180523	9	CNS01DSH Human chr
C	28	549	86.5	190565	9	AC007374 Homo sapi
C	29	549	86.5	203777	9	CNS01DVH Human chr
C	30	547.8	86.3	112544	2	AC009443 Homo sapi
C	31	547.6	86.2	192178	2	AC009443 Homo sapi
C	32	547.6	86.2	192899	9	CNS05TFS Human chr
C	33	547.4	86.2	172281	9	AC068492 Homo sapi
C	34	547.2	86.2	169418	9	AC104163 Homo sapi
C	35	546.6	86.1	161054	9	AC104444 Homo sapi
C	36	546.4	86.0	2074	6	AX478550 Sequence
C	37	546	86.0	169234	9	CNS06C7Q Homo sapi
C	38	545	85.8	1329	6	AR344389 Sequence
C	39	545	85.8	1329	6	AX001030 Sequence
C	40	545	85.8	1329	6	BD136199 Retrovira
C	41	545	85.8	99408	9	AL359385 Human DNA
C	42	545	85.8	170586	2	AL356126 Homo sapi
C	43	543.8	85.6	181166	9	AC087897 Homo sapi
C	44	543.4	85.6	88328	9	AL357874 Human DNA
C	45	539	84.9	2082	2	AC134379 Homo sapi

ALIGNMENTS

RESULT 1	AR344385	Sequence 102 from patent US 6582703.	635 bp	mRNA	linear	PAT 17-AUG-2003
LOCUS	AR344385					
DEFINITION	AR344385					
ACCESSION	AR344385.1	GI:33740326				
VERSION						
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unclassified.					
REFERENCE	1 (bases 1 to 635)					
AUTHORS	Perron, H., Beseme, F., Bedin, F., Paranhos-Baccala, G., Komurian-Pradel, F., Jolivet-Reynaud, C. and Mandrand, B.					
TITLE	Isolated nucleotide sequences associated with multiple sclerosis or rheumatoid arthritis and a process of detecting					
JOURNAL	Patent: US 6582703-A 102 24-JUN-2003;					
FEATURES	Location/Qualifiers					
source	1. .635					
	/organism="unknown"					
	/mol_type="mRNA"					

ORIGIN

Query Match	100.0%;	Score 635;	DB 6;	Length 635;
Best Local Similarity	100.0%;	Pred. No. 1.1e-189;		
Matches 635;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	CCCTGATCTTTAACTCTCTTTGTTAGTTTCTCTTCCAGATCAAACTGTAATACTA 60		
Db	1	CCCTGATCTTTAACTCTCTTTGTTAGTTTCTCTTCCAGATCAAACTGTAATACTA 60		
Qy	61	CAAATTTGTTCTTCAATGGAGCACCAGATGAGTCCATGACTTAAGATCCACCGTGGACCC 120		
Db	61	CAAATTTGTTCTTCAATGGAGCACCAGATGAGTCCATGACTTAAGATCCACCGTGGACCC 120		
Qy	121	CTGGACCGGCTGCTAGCCCATCTCCGATGTTTAATGACATTAAGAGCACCCTCCCGAG 180		
Db	121	CTGGACCGGCTGCTAGCCCATCTCCGATGTTTAATGACATTAAGAGCACCCTCCCGAG 180		
Qy	181	GAATCTCACTGCACAAACCCCTACTATGCCCAATTCAGCGGAGGAGCTTACAGCGGT 240		
Db	181	GAATCTCACTGCACAAACCCCTACTATGCCCAATTCAGCGGAGGAGCTTACAGCGGT 240		

```
QY 241 CATACGCCAACCTCCCAACAGACACTTGGGTTTTCTGTGTGAGAGGGGGGACTGAGAGAC 300
Db 241 CATACGCCAACCTCCCAACAGACACTTGGGTTTTCTGTGTGAGAGGGGGGACTGAGAGAC 300
QY 301 AGGACTAGCTGGATTTCTTAGGCCAACAGAGAAATCCCTAAGCCTAGCTGGGAGGTGACT 360
Db 301 AGGACTAGCTGGATTTCTTAGGCCAACAGAGAAATCCCTAAGCCTAGCTGGGAGGTGACT 360
QY 361 GCATCCACCTCTAAACATGCGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
Db 361 GCATCCACCTCTAAACATGCGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
QY 421 ACTAAATGCTAATTAGGCAAAAATAGGAGGTAAGAAATAGCAATCATCTATTTCCTG 480
Db 421 ACTAAATGCTAATTAGGCAAAAATAGGAGGTAAGAAATAGCAATCATCTATTTCCTG 480
QY 481 AGAGCACGCGGAGGACAGAGATCGGATATAAACCCAGGCAATTCGAGCGGCAACGG 540
Db 481 AGAGCACGCGGAGGACAGAGATCGGATATAAACCCAGGCAATTCGAGCGGCAACGG 540
QY 541 CAACCCCTTTGGGTCCTCCCTTTGTATGGCGCTCTGTGTTTCACTCTATTTCCTCT 600
Db 541 CAACCCCTTTGGGTCCTCCCTTTGTATGGCGCTCTGTGTTTCACTCTATTTCCTCT 600
QY 601 ATTAATCTTGCACCTGAAAAAAGAAAAA 635
Db 601 ATTAATCTTGCACCTGAAAAAAGAAAAA 635
```

RESULT 2

```
AX001024
LOCUS AX001024
DEFINITION Sequence 6 from Patent WO9902666.
ACCESSION AX001024
VERSION AX001024.1 GI:7241262
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 635)
AUTHORS Ott,C. and Bedin,F.
TITLE RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES
JOURNAL Patent: WO 9902666-A 6 21-JAN-1999;
BIO MERIEUX (FR); OTT CATHERINE (FR)
FEATURES
source
1. .635
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
```

ORIGIN

```
Query Match 100.0%; Score 635; DB 6; Length 635;
Best Local Similarity 100.0%; Pred. No. 1.le-189;
Matches 635; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCCTGTATCTTTAACTCCCTTGTAGTTTGTCTTCCAGAAATCAAACTGTAATACTA 60
Db 1 CCCTGTATCTTTAACTCCCTTGTAGTTTGTCTTCCAGAAATCAAACTGTAATACTA 60
QY 61 CAAATTTCTTCAAATGGAGCACAGATGGATGCCATGATCAAGATCCACCGTGGACCC 120
Db 61 CAAATTTCTTCAAATGGAGCACAGATGGATGCCATGATCAAGATCCACCGTGGACCC 120
QY 121 CTGACCGCGCTGTAGCCCATGCTCCGATGTTAATGACATTTGAGGACCCCTCCCGAG 180
Db 121 CTGACCGCGCTGTAGCCCATGCTCCGATGTTAATGACATTTGAGGACCCCTCCCGAG 180
QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAACAGTTAGAGCGGT 240
Db 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAACAGTTAGAGCGGT 240
```

```
QY 241 CATACGCCAACCTCCCAACAGACACTTGGGTTTTCTGTGTGAGAGGGGGGACTGAGAGAC 300
Db 241 CATACGCCAACCTCCCAACAGACACTTGGGTTTTCTGTGTGAGAGGGGGGACTGAGAGAC 300
QY 301 AGGACTAGCTGGATTTCTTAGGCCAACAGAGAAATCCCTAAGCCTAGCTGGGAGGTGACT 360
Db 301 AGGACTAGCTGGATTTCTTAGGCCAACAGAGAAATCCCTAAGCCTAGCTGGGAGGTGACT 360
QY 361 GCATCCACCTCTAAACATGCGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
Db 361 GCATCCACCTCTAAACATGCGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
QY 421 ACTAAATGCTAATTAGGCAAAAATAGGAGGTAAGAAATAGCAATCATCTATTTCCTG 480
Db 421 ACTAAATGCTAATTAGGCAAAAATAGGAGGTAAGAAATAGCAATCATCTATTTCCTG 480
QY 481 AGAGCACGCGGAGGACAGAGATCGGATATAAACCCAGGCAATTCGAGCGGCAACGG 540
Db 481 AGAGCACGCGGAGGACAGAGATCGGATATAAACCCAGGCAATTCGAGCGGCAACGG 540
QY 541 CAACCCCTTTGGGTCCTCCCTTTGTATGGCGCTCTGTGTTTCACTCTATTTCCTCT 600
Db 541 CAACCCCTTTGGGTCCTCCCTTTGTATGGCGCTCTGTGTTTCACTCTATTTCCTCT 600
QY 601 ATTAATCTTGCACCTGAAAAAAGAAAAA 635
Db 601 ATTAATCTTGCACCTGAAAAAAGAAAAA 635
```

RESULT 3

```
BD136195
LOCUS BD136195
DEFINITION Retroviral nucleic material and nucleotide fragments, in particular, associated with multiple sclerosis and/or rheumatoid arthritis, for diagnostic, prophylactic and therapeutic uses.
ACCESSION BD136195
VERSION BD136195.1 GI:23231140
KEYWORDS JP 2002509437-A/5.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 635)
AUTHORS Baccala,G.P., Pradel,F.K., Bedin,F., Sodoyer,M., Ott,C., Mallet,F., Perron,H. and Mandrand,B.
TITLE Retroviral nucleic material and nucleotide fragments, in particular, associated with multiple sclerosis and/or rheumatoid arthritis, for diagnostic, prophylactic and therapeutic uses
JOURNAL Patent: JP 2002509437-A 5 26-MAR-2002;
BIO MERIEUX
COMMENT
OS Unidentified
PN JP 2002509437-A/5
PD 26-MAR-2002
PE 07-JUL-1998 JP 1999508255
PR 07-JUL-1997 FR 97/08816
PI GLAUCIA PARAINOS BACCALA, FLORENCE KOMURIAN PRADEL, FREDERIC PI
BEDIN,
PI MIREILLE SODOYER, CATHERINE OTT, FRANCOIS MALLET, HERVE PERRON,
PI BERNARD MANDRAND
PC C12N15/48, C12Q1/70, C07K14/15, A61K31/70
CC Strandedness: Single;
CC Topology: Linear;
CC Retroviral nucleic material and nucleotide fragments, in CC
CC associated with multiple sclerosis and/or
rheumatoid arthritis,
for
CC diagnostic, prophylactic and therapeutic uses PH Key
CC Location/Qualifiers
FT source
1. .635
/organism="unidentified".
FT Location/Qualifiers
1. .635
/organism="unidentified"
```

```
ORIGIN
/mol_type="genomic DNA"
/db_xref="taxon:32644"

Query Match      100.0%; Score 635; DB 6; Length 635;
Best Local Similarity 100.0%; Pred. No. 1.1e-189;
Matches 635; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTGTATCTTTAACTCTCTGTTAAAGTTTGTCTTCCAGAAATCAAAACTGTAAACTA 60
DB 1 CCTGTATCTTTAACTCTCTGTTAAAGTTTGTCTTCCAGAAATCAAAACTGTAAACTA 60
QY 61 CAAATTTGTTCTTCAATGGAGCACCAGATGGAGTCCATGATGAATCAAGTCCACCGTGGACCC 120
DB 61 CAAATTTGTTCTTCAATGGAGCACCAGATGGAGTCCATGATGAATCAAGTCCACCGTGGACCC 120
QY 121 CTGGACCGGCTCTGATGCCCATGCTCCGATGTTAAATGACATTAAGAGCCACCCCTCCCGAG 180
DB 121 CTGGACCGGCTCTGATGCCCATGCTCCGATGTTAAATGACATTAAGAGCCACCCCTCCCGAG 180
QY 181 GAAATCTCAACTGCAACAACCCCTACTATGCCCCCAATTCAGCGGGAAGCAGTTAGACGGGT 240
DB 181 GAAATCTCAACTGCAACAACCCCTACTATGCCCCCAATTCAGCGGGAAGCAGTTAGACGGGT 240
QY 241 CATCAGCAACCTCCCAACAGCACTTGGGTTTTCTGTTGAGAGGGGGGACTGAGAGAC 300
DB 241 CATCAGCAACCTCCCAACAGCACTTGGGTTTTCTGTTGAGAGGGGGGACTGAGAGAC 300
QY 301 AGGACTAGCTGGATTTCTTAGGCCAACAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACT 360
DB 301 AGGACTAGCTGGATTTCTTAGGCCAACAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACT 360
QY 361 GCATCCACCTCTAAACATGCGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
DB 361 GCATCCACCTCTAAACATGCGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
QY 421 ACTAAATGCTAATTAGGCAAAATAGAGGTAAAGAAATAGCAATATCAATCATCTATTCCTG 480
DB 421 ACTAAATGCTAATTAGGCAAAATAGAGGTAAAGAAATAGCAATATCAATCATCTATTCCTG 480
QY 481 AGAGCACAGCGGGAGGACAGGATCGGGATATAAACCCAGGATTCGAGCGGGAACGG 540
DB 481 AGAGCACAGCGGGAGGACAGGATCGGGATATAAACCCAGGATTCGAGCGGGAACGG 540
QY 541 CAACCCCTTTGGGTCCTCCCTCTGTTGATGGCGCTCTGTTTCACTCTATTTCACTCT 600
DB 541 CAACCCCTTTGGGTCCTCCCTCTGTTGATGGCGCTCTGTTTCACTCTATTTCACTCT 600
QY 601 ATTAATCTTGCAACTGAAAAAAGAAAAA 635
DB 601 ATTAATCTTGCAACTGAAAAAAGAAAAA 635

RESULT 4
AF127229 2004 bp mRNA linear VRL 11-AUG-1999
LOCUS Multiple sclerosis associated retrovirus element clone CL6
DEFINITION pol-env/3'LTR-like mRNA sequence.
ACCESSION AF127229
VERSION AF127229.1 GI:5726294
KEYWORDS Multiple sclerosis associated retrovirus element
SOURCE Multiple sclerosis associated retrovirus element
ORGANISM Viruses; Retroviridae.
REFERENCE 1 (bases 1 to 2004)
AUTHORS Komurian-Pradel,F., Paranhos-Baccala,G., Bedin,F.,
Ounanian-Paraz,A., Sodoyer,M., Ott,C., Rajoharison,A., Garcia,E.,
Mallet,F., Mandrand,B. and Perron,H.
TITLE Molecular cloning and characterization of MSRV-related sequences
JOURNAL Virology 260 (1), 1-9 (1999)
MEDLINE 99335590
PUBMED 10405350

REFERENCE 2 (bases 1 to 2004)
AUTHORS Komurian-Pradel,F., Paranhos-Baccala,G., Bedin,F.,
Ounanian-Paraz,A., Sodoyer,M., Ott,C., Rajoharison,A., Garcia,E.,
Mallet,F., Mandrand,B. and Perron,H.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-1999) UMR103 CNRS, bioMerieux, 46, allée
D'Italie, Lyon 69007, France
FEATURES
source
1..2004
Location/Qualifiers
/organism="Multiple sclerosis associated retrovirus
element"
/mol_type="mRNA"
/db_xref="taxon:89382"
/clone="CL6"
1..2004
misc_feature
/note="similar to pol-env and 3'LTR region"

ORIGIN
Query Match      100.0%; Score 635; DB 14; Length 2004;
Best Local Similarity 100.0%; Pred. No. 1.3e-189;
Matches 635; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTGTATCTTTAACTCTCTGTTAAAGTTTGTCTTCCAGAAATCAAAACTGTAAACTA 60
DB 1370 CCTGTATCTTTAACTCTCTGTTAAAGTTTGTCTTCCAGAAATCAAAACTGTAAACTA 1429
QY 61 CAAATTTGTTCTTCAATGGAGCACCAGATGGAGTCCATGATGAATCAAGTCCACCGTGGACCC 120
DB 1430 CAAATTTGTTCTTCAATGGAGCACCAGATGGAGTCCATGATGAATCAAGTCCACCGTGGACCC 1489
QY 121 CTGGACCGGCTCTGATGCCCATGCTCCGATGTTAAATGACATTAAGAGCCACCCCTCCCGAG 180
DB 1490 CTGGACCGGCTCTGATGCCCATGCTCCGATGTTAAATGACATTAAGAGCCACCCCTCCCGAG 1549
QY 181 GAAATCTCAACTGCAACAACCCCTACTATGCCCCCAATTCAGCGGGAAGCAGTTAGACGGGT 240
DB 1550 GAAATCTCAACTGCAACAACCCCTACTATGCCCCCAATTCAGCGGGAAGCAGTTAGACGGGT 1609
QY 241 CATCAGCAACCTCCCAACAGCACTTGGGTTTTCTGTTGAGAGGGGGGACTGAGAGAC 300
DB 1610 CATCAGCAACCTCCCAACAGCACTTGGGTTTTCTGTTGAGAGGGGGGACTGAGAGAC 1669
QY 301 AGGACTAGCTGGATTTCTTAGGCCAACAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACT 360
DB 1670 AGGACTAGCTGGATTTCTTAGGCCAACAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACT 1729
QY 361 GCATCCACCTCTAAACATGCGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
DB 1730 GCATCCACCTCTAAACATGCGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 1789
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATATCAATCATCTATTCCTG 480
DB 1790 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATATCAATCATCTATTCCTG 1849
QY 481 AGAGCACAGCGGGAGGACAGGATCGGGATATAAACCCAGGATTCGAGCGGGAACGG 540
DB 1850 AGAGCACAGCGGGAGGACAGGATCGGGATATAAACCCAGGATTCGAGCGGGAACGG 1909
QY 541 CAACCCCTTTGGGTCCTCCCTCTGTTGATGGCGCTCTGTTTCACTCTATTTCACTCT 600
DB 1910 CAACCCCTTTGGGTCCTCCCTCTGTTGATGGCGCTCTGTTTCACTCTATTTCACTCT 1969
QY 601 ATTAATCTTGCAACTGAAAAAAGAAAAA 635
DB 1970 ATTAATCTTGCAACTGAAAAAAGAAAAA 2004

RESULT 5
AF113137/c 167694 bp DNA linear PRI 25-APR-2002
LOCUS Homo sapiens chromosome 18, clone RP11-622J9, complete sequence.
DEFINITION AC113137
ACCESSION AC113137.2 GI:19683504
VERSION
```

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 167694)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.

JOURNAL

Unpublished
2 (bases 1 to 167694)

AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fardo,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

JOURNAL

Submitted (25-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

3 (bases 1 to 167694)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Fardo,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

JOURNAL

Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

4 (bases 1 to 167694)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Fardo,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,

McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

JOURNAL

Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Mar 22, 2002 this sequence version replaced gi:118875226.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L24497

Center clone name: 622_J_9

FEATURES

source

1..167694
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-622J9"
/clone_lib="RPC1-11 Human Male BAC"
2566..2593
/rpt_family="AT_rich"
3108..3127
/rpt_family="(TA)n"
3128..3153
/rpt_family="(TG)n"
3154..3208
/rpt_family="(TA)n"
complement(3212..3818)
/rpt_family="LIMA9"
4840..5079
/rpt_family="LIME3"
5234..6028
/rpt_family="LIME3"
complement(6259..6470)
/rpt_family="MIR"
6723..6750
/rpt_family="AT_rich"
6982..7280
/rpt_family="AluY"
7780..7811
/rpt_family="AT_rich"
complement(8120..8302)
/rpt_family="LIM4"
complement(8444..8971)
/rpt_family="LIME3A"
complement(8973..9386)
/rpt_family="LIME3A"
complement(9387..9690)
/rpt_family="LIMA9"
9898..10497
/rpt_family="MLTIE3"
11218..11555
/rpt_family="L2"
complement(12442..13393)
/rpt_family="LIPA16"
13550..13907

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region


```

----- Project Information
Center project name: L4296
Center clone name: 497_M_7
----- Summary Statistics
Sequencing vector: M13; M77815; 32% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 176087 bases at least Q40
Consensus quality: 176095 bases at least Q30
Consensus quality: 176095 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 176095; sum-of-contigs
Quality coverage: 11.5 in Q20 bases; agarose-fp
Quality coverage: 11.7 in Q20.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 176095: contig of 176095 bp in length.
assembly_fragment.
FEATURES             Location/Qualifiers
     source            1..176095
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="18"
                        /map="18"
                        /clone="RP11-497M7"
                        /clone_lib="RPC1-11 Human Male BAC"

ORIGIN
Query Match          99.5%; Score 631.8; DB 2; Length 176095;
Best Local Similarity 99.7%; Pred. No. 2.9e-188;
Matches 633; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCCTGTATCTTTAACTCCTTGTAGTTTGTCTCTCCAGATCAAACTGTAAACTA 60
Db 150524 CCCTGTATCTTTAACTCCTTGTAGTTTGTCTCTCCAGATCAAAACCGTAA 150465

Qy 61 CAAATGTTCTTCAAATGGAGACCCAGATGGATCCATGACTAAGATCCACCGTGACCC 120
Db 150464 CAAATGTTCTTCAAATGGAGACCCAGATGGATCCATGACTAAGATCCACCGTGACCC 150405

Qy 121 CTGACCGGCGCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
Db 150404 CTGACCGGCGCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 150345

Qy 181 GAAATCTCAACTGACAAACCCCTACTATGCCCAATTCAGCGGAGCAGTTAGAGCGGT 240
Db 150344 GAAATCTCAACTGACAAACCCCTACTATGCCCAATTCAGCGGAGCAGTTAGAGCGGT 150285

Qy 241 CATCAGCAACCTCCCAACAGACATTGGGTTTCTGTTGAGAGGGGGAGCTGAGAGAC 300
Db 150284 CATCAGCAACCTCCCAACAGACATTGGGTTTCTGTTGAGAGGGGGAGCTGAGAGAC 150225

Qy 301 AGGACTAGCTGGATTTCTTAGGCCCAAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACT 360
Db 150224 AGGACTAGCTGGATTTCTTAGGCCCAAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACT 150165

Qy 361 GCATCCACCTTAAACATGGGGCTTGCACTTAGCTCACAACCCGACCAATCAGAGCTC 420
Db 150164 GCATCCACCTTAAACATGGGGCTTGCACTTAGCTCACAACCCGACCAATCAGAGCTC 150105

Qy 421 ACTAAATGCTAATTAGGCAAAATAGGAGTAAAGAAATAGCAATCATCTATTGCTG 480
Db 150104 ACTAAATGCTAATTAGGCAAAATAGGAGTAAAGAAATAGCAATCATCTATTGCTG 150045

```

```

QY 481 AGAGCACAGCGGAGGACAGGATCGGATATAAACCCAGGATTCGAGCGGCAACGG 540
Db 150044 AGAGCACAGCGGAGGACAGGATCGGATATAAACCCAGGATTCGAGCGGCAACGG 149985

QY 541 CAACCCCTTTGGGTCCCTTCCCTTTGTATGGCGCTCTGTTTTCATCTATTTCATCT 600
Db 149984 CAACCCCTTTGGGTCCCTTCCCTTTGTATGGCGCTCTGTTTTCATCTATTTCATCT 149925

QY 601 ATTAATCTTGCAACTGAAAAAAGAAAAAAGAAAAA 635
Db 149924 ATTAATCTTGCAACTGAAAAAAGAAAAAAGAAAAA 149890

RESULT 7
AC022171 210336 bp DNA linear HTG 26-MAR-2001
LOCUS Homo sapiens chromosome 18 clone RP11-407C18, WORKING DRAFT
DEFINITION SEQUENCE, 2 unordered pieces.
AC022171
AC022171.18 GI:13270574
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 210336)
Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
Unpublished
2 (bases 1 to 210336)
Bruno,D., Conn,L., Dela Rosa,M., Federspiel,N., Foreman,P.,
Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,
Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D., Wilhelmy,J.,
Yu,S. and Davis,R.W.
Direct Submission
Submitted (26-JAN-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
On Mar 10, 2001 this sequence version replaced gi:13122770.
----- Genome Center
Center: Stanford DNA Sequencing and Technology Development
Center
Center code: SDSTDC
Web site: http://sequence-www.stanford.edu/group/human/
Contact: hum-info@sequence.stanford.edu
----- Project Information
Center project name: 758
Center clone name: RP11-407C18
----- Summary Statistics
Sequencing Vector: M13mp18; X02513; 97% of reads
Sequencing Vector: plasmid; plasmid_accession; 3% of reads
Chemistry: Dye-terminator Big Dye; 10% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 208312 bases at least Q40
Consensus quality: 208877 bases at least Q30
Consensus quality: 209157 bases at least Q20
Insert size: 199762; agarose-fp
Insert size: 210236; sum-of-contigs
Quality coverage: 8.7x in Q20 bases; agarose-fp
Quality coverage: 8.3x in Q20 bases; sum-of-contigs.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 66468: contig of 66468 bp in length

```

```
* 66469 66568: gap of unknown length
* 66569 210336: contig of 143768 bp in length.
FEATURES
  source      Location/Qualifiers
    1..210336
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="18"
      /clone="RP11-407C18"
      /clone_lib="RP11-407C18"
      /note="assembly_name:Contig44"
  misc_feature 1..66468
  misc_feature 66569..210336
      /note="assembly_name:Contig45"
      /clone_end:SP6"
ORIGIN
  Query Match      99.5%; Score 631.8; DB 2; Length 210336;
  Best Local Similarity 99.7%; Pred. No. 3e-186;
  Matches 633; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
  QY 1 CCTGTATCTTTAACTCCTCTTGAAGTTTGTCTCTTCCAGAAATCAAACTGTAATAACTA 60
  Db 58176 CCTGTATCTTTAACTCCTCTTGAAGTTTGTCTCTTCCAGAAATCAAACTGTAATAACTA 58235
  QY 61 CAAATGTTTCTTCAATGGAGCACCAGATGAGTCCATGATCAAGATCCACCGTGGACCC 120
  Db 58236 CAAATGTTTCTTCAATGGAGCACCAGATGAGTCCATGATCAAGATCCACCGTGGACCC 58295
  QY 121 CTGGACGGCTCTAGCCCATCTCGATGTTAATGACATTTGAAGCACCCTCCCGAG 180
  Db 58296 CTGGACGGCTCTAGCCCATCTCGATGTTAATGACATTTGAAGCACCCTCCCGAG 58355
  QY 181 GAAATCTCAACTGCACAAACCTTACTATGCCCAATTCAGCGGAAGAGTGTAGACGGT 240
  Db 58356 GAAATCTCAACTGCACAAACCTTACTATGCCCAATTCAGCGGAAGAGTGTAGACGGT 58415
  QY 241 CATCAGCCAACTCCCAACAGACATTTGGTGTCTTGTGAGAGGGGGAGTGTAGAGAC 300
  Db 58416 CATCAGCCAACTCCCAACAGACATTTGGTGTCTTGTGAGAGGGGGAGTGTAGAGAC 58475
  QY 301 AGACTAGCTGGATTTCTTAGGCCAACGAAAGATCCCTAGCTAGCTAGCTGGAGGTGACT 360
  Db 58476 AGACTAGCTGGATTTCTTAGGCCAACGAAAGATCCCTAGCTAGCTAGCTGGAGGTGACT 58535
  QY 361 GCATCCACCTCTAAACATGGGCTTGCACCTTAGCTCACACCCGCAATTCAGAGAGCTC 420
  Db 58536 GCATCCACCTCTAAACATGGGCTTGCACCTTAGCTCACACCCGCAATTCAGAGAGCTC 58595
  QY 421 ACTAAATGCTATTAGGCAAAATAGGAGGTAAGAAATAGCAATCATCTATTGCTG 480
  Db 58596 ACTAAATGCTATTAGGCAAAATAGGAGGTAAGAAATAGCAATCATCTATTGCTG 58655
  QY 481 AGAGCAGCGGAGGAGCAAGATCGGATATAAACCCAGGCAATTCGAGCCGGCAACGG 540
  Db 58656 AGAGCAGCGGAGGAGCAAGATCGGATATAAACCCAGGCAATTCGAGCCGGCAACGG 58715
  QY 541 CAACCCCTTTGGGTCCCTCTTGTATGGGCTCTGTTTTCACCTCTATTTCACCTCT 600
  Db 58716 CAACCCCTTTGGGTCCCTCTTGTATGGGCTCTGTTTTCACCTCTATTTCACCTCT 58775
  QY 601 ATTAATCTTGCACCTGAAAAAAGAAAAAAGAAAAA 635
  Db 58776 ATTAATCTTGCACCTGAAAAAAGAAAAAAGAAAAA 58810
  RESULT 8
  AC010778/c 148724 bp DNA linear HTG 30-MAR-2000
  LOCUS
  DEFINITION Homo sapiens clone RP11-2N15, WORKING DRAFT SEQUENCE, 8 unordered
  pieces.
  AC010778
  AC010778
```

VERSION
KEYWORDS
SOURCE
ORGANISM

AC010778.3 GI:7341824
HTG; HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens (human)
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 148724)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Conell,L., Doyle,M.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donnell,P.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,N., Klein,J.,
Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,A.,
McSwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 30, 2000 this sequence version replaced gi:6087973.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2677
Center clone name: 2 N 15
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 140899 bases at least Q40
Consensus quality: 144351 bases at least Q30
Consensus quality: 145949 bases at least Q20
Insert size: 182000; agarose-fp
Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 4.8 in Q20 bases; sum-of-contigs

TITLE
JOURNAL

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2462: contig of 2462 bp in length
* 2463 2562: gap of 100 bp
* 2563 7948: contig of 5386 bp in length
* 7949 8048: gap of 100 bp
* 8049 15800: contig of 7752 bp in length
* 15801 15900: gap of 100 bp
* 15901 25059: contig of 9159 bp in length
* 25060 41906: contig of 16747 bp in length
* 41907 42006: gap of 100 bp
* 42007 63441: contig of 21335 bp in length
* 63442 98587: contig of 35146 bp in length

QY	601	ATTAATCTTGCRAACTGAAAAA 633	RESULT 9	AL162912
Db	39671	ATTAATCTTGCRAACTGAAAAAATGC 39639	LOCUS	AL162912
			DEFINITION	Human DNA sequence from clone RP6-198C4 on chromosome Xq13.3-21.2, complete sequence.
			ACCESSION	AL162912
			VERSION	AL162912.1
			KEYWORDS	HTG.
			SOURCE	Homo sapiens (human)
			ORGANISM	Homo sapiens
			REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
			AUTHORS	1 (bases 1 to 8339)
			TITLE	Pearce, A.
			JOURNAL	Direct Submission
			COMMENT	Submitted (22-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/ChrX This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP6-198C4 is from the library RPCI-6 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pPAC4
			FEATURES	Location/Qualifiers
			source	1. .8339
				/organism="Homo sapiens"
				/mol_type="genomic DNA"
				/db_xref="taxon:9606"
				/chromosome="X"
				/map="q13.3-21.2"
				/clone="RP6-198C4"
				/clone_lib="RPCI-6"
			repeat_region	1. .5641
				/note="HERV17 repeat: matches 368. .5930 of consensus"
			repeat_region	5635. .5673
				/note="HERV17 repeat: matches 7419. .7457 of consensus"
			repeat_region	5664. .5955
				/note="HERV17 repeat: matches 8244. .8523 of consensus"
			repeat_region	5956. .6136
				/note="LTR17 repeat: matches 1. .239 of consensus"
			repeat_region	6195. .6280
				/note="LTR17 repeat: matches 239. .324 of consensus"
			ORIGIN	
			Query Match	91.4%; Score 580.2; DB 2; Length 148724;
			Best Local Similarity	94.8%; Pred. No. 6.7e-172;
			Matches 600; Conservative	0; Mismatches 33; Indels 0; Gaps 0;
QY	1	CCCTGTATCTTTAACTCCCTGTAGTTGCTCTTCCAGATCAAACTGTA 60		
Db	40271	CCCTGTATCTTTAACTCCCTGTAGTTGCTCTTCCAGATTTGAAC 40212		
QY	61	CAAAATGTTCTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120		
Db	40211	CAAAATGTTCTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGGATCC 40152		
QY	121	CTGACCGGCTGTAGCCCATGCTCCGATGTTAATGAATGAAGCACCCTCCCGAG 180		
Db	40151	CTGACCGGCTGTAGCCCATGCTCCGATGTTAATGGCATCGAAGGACCCCTCCCGAG 40092		
QY	181	GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAACGAGTGGAGCGGT 240		
Db	40091	GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAACGAGTGGAGCGGT 40032		
QY	241	CATCAGCCAACTCCCAACAGCACTTGGGTTTTCTGTTGAGAGGGGAGTGGAGAC 300		
Db	40031	CATCAGCCAACTCCCAACAGCACTTGGGTTTTCTGTTGAGAGGGGAGTGGAGAC 39972		
QY	301	AGGACTAGTGGATTCCTTAGGCCAACAGAGATCCCTAGCTAGCTGGAGAGTGA 360		
Db	39971	AGGACTAGTGGATTCCTTAGGCCAACAGAGATCCCTAGCTAGCTGGAGAGTGA 39912		
QY	361	GCATCCACCTCTAAACATGGGCTTGAATAGCTTACACCCGACCAATCAGAGAGCTC 420		
Db	39911	GCATCCACCTCTAAACATGGGCTTGAATAGCTTACACCCGACCAATCAGAGAGCTC 39852		
QY	421	ACTAAATGCTTAATAGGCAAAATAGGAGTAAAGAAATAGCAATCATCTATTGCTG 480		
Db	39851	ACTAAATGCTTAATAGGCAAAATAGGAGTAAAGAAATAGCAATCATCTATTGCTG 39792		
QY	481	AGAGCAGCGGAGGAGCAAGATTCGGATATTAACCCAGGCAATTCAGCCGCAACGG 540		
Db	39791	AGAGCAGCGGAGGAGCAAGATTCGGATATTAACCCAGGCAATTCAGCCGCAACAG 39732		
QY	541	CAACCCCTTTGGTCCCTCCCTCTTGTATGGGGCTCTGTTTCACTCTATTTCAC 600		
Db	39731	CAACCCCTTTGGTCCCTCCCTCTTGTATGGGGCTCTGTTTCACTCTATTTCAC 39672		

QY	601	ATTAATCTTGCACTGAAAAA 533	QY	601	ATTAATCTTGCACTGAAAAA 533
Db	39671	ATTAATCTTGCACTGAAAAAATGC 39639	Db	39671	ATTAATCTTGCACTGAAAAAATGC 39639
RESULT 9					
AL162912			AL162912		
LOCUS			LOCUS		
DEFINITION			DEFINITION		
Human DNA sequence from clone RP6-198C4 on chromosome Xq13.3-21.2, complete sequence.			Human DNA sequence from clone RP6-198C4 on chromosome Xq13.3-21.2, complete sequence.		
ACCESSION			ACCESSION		
AL162912			AL162912		
VERSION			VERSION		
AL162912.1			AL162912.1		
KEYWORDS			KEYWORDS		
HTG.			HTG.		
SOURCE			SOURCE		
Homo sapiens (human)			Homo sapiens (human)		
ORGANISM			ORGANISM		
Homo sapiens			Homo sapiens		
REFERENCE			REFERENCE		
1 (bases 1 to 8339)			1 (bases 1 to 8339)		
Direct Submission			Direct Submission		
Submitted (22-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk			Submitted (22-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk		
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.			During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.		
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at			The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at		
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at			http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at		
http://www.sanger.ac.uk/HGP/ChrX			http://www.sanger.ac.uk/HGP/ChrX		
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RF6-198C4 is from the library RPC1-6 constructed by the group of Pieter de Jong. For further details see			This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RF6-198C4 is from the library RPC1-6 constructed by the group of Pieter de Jong. For further details see		
http://www.chori.org/bacpac/home.htm			http://www.chori.org/bacpac/home.htm		
VECTOR: pPAC4			VECTOR: pPAC4		
IMPORTANT: This sequence is not the entire insert of clone RP6-198C4. It may be shorter because we sequence overlapping sections only once, except for a short overlap.			IMPORTANT: This sequence is not the entire insert of clone RP6-198C4. It may be shorter because we sequence overlapping sections only once, except for a short overlap.		
FEATURES					
Location/Qualifiers					
1. .8339					
/organism="Homo sapiens"					
/mol_type="genomic DNA"					
/db_xref="taxon:9606"					
/chromosome="X"					
/map="q13.3-21.2"					
/clone="RP6-198C4"					
/clone_lib="RPC1-6"					
1. .5641					
/notes="HERV17 repeat: matches 368. .5930 of consensus"					
repeat_region					
5635. .5673					
/notes="HERV17 repeat: matches 7419. .7457 of consensus"					
repeat_region					
5664. .5955					
/notes="HERV17 repeat: matches 8244. .8523 of consensus"					
repeat_region					
5956. .6136					
/notes="LTR17 repeat: matches 1. .239 of consensus"					
repeat_region					
6195. .6280					
/notes="LTR17 repeat: matches 239. .324 of consensus"					
repeat_region					
ORIGIN					

```

Query Match      91.2%  Score 579.2;  DB 9;  Length 8339;
Best Local Similarity 94.8%;  Pred. No. 8.4e-172;  Indels 0;  Gaps 0;
Matches 599;  Conservative 0;  Mismatches 33;

QY 1 CCTGTATCTTTAACTCCCTTGTGTTAGTTGTTCTCTCCAGATCAAACTGTAATACTA 60
Db 5664 CCCTGTATCTTTAACTCCCTTGTGTTAGTTGTTCTCTCCAGATTTGAACTGTAATACTA 5723

QY 61 CAAATGTTCTTCAATGGAGCACCAGATGAGTCCATGATCAAGATCCACCGTGACCC 120
Db 5724 CAAATGTTCTTCAATGGAGCACCAGATGAGTCCATGATCAAGATCCACCGTGATCC 5783

QY 121 CTGACCGGCTGCTAGCCCATGCTCCGATGTTAATGACATTTGAAGCACCCTCCCGAG 180
Db 5784 CTGACCGGCTGCTAGCCCATGCTCCGATGTTAATGATGCAAGCACCCTCCCGAG 5843

QY 181 GAAATCTCAACTGACCAACCCCTACTATGCCCCAAATTCAGCGGGAAGCAGTTAGACGGT 240
Db 5844 GAAATCTCAACTGACCAACCCCTACTATGCCCCAAATTCAGCAGGAGCAGTTAGACGAGT 5903

QY 241 CATCAGCAACCTCCCAACAGACATTTGGTTTCTGTTGAGAGGGGAGCTGAGAGAC 300
Db 5904 CATCAGCAACCTCCCAACAGACATTTGGTTTCTGTTGAGAGGGGAGCTGAGAGAC 5963

QY 301 AGCACTAGCTGGATTTCTAGGCCCAAGCAAGATCCCTAGACCTAGCTGGAGAGTGACT 360
Db 5964 AGCACTAGCTGGATTTCTAGGCCCAAGCAAGATCCCTAGACCTAGCTGGAGAGTGACT 6023

QY 361 GCATCACCCTTAAACATGCGGGCTTGCAACTTACCTACACACCCGACCAATCAGAGAGCTC 420
Db 6024 ACACCCACCTTTAAACATGCGGGCTTGCAACTTACCTACACACCCGACCAATCAGAGAGCTC 6083

QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAGAAATPAGCAATCATCTATTGCGCTG 480
Db 6084 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAGAAATPAGCAATCATCTATTGCGCTG 6143

QY 481 AGAGCAGCGGAGGAGGAGCAGATCGGATATTAACCCAGGATTCGAGCCGGCAACGG 540
Db 6144 AGAGCAGCGTGGAGGAGCAGATCGGATATTAACCCAGGATTCGAGCCGGCAACAG 6203

QY 541 CAACCCCTTTGGGTCCTCCCTCTGTTATGGGCTCTGTTTCACTCTATTTCCTCT 600
Db 6204 CAACCCCTTTGGGTCCTCCCTCTGTTATGGGCTCTGTTTCACTCTATTTCCTCT 6263

QY 601 ATTAATCTTGCAACTGAAAAAATAAAAAA 632
Db 6264 ATTAATCTTGCAACTGAAAAAATAAAAAA 6295

RESULT 10
AC073626/c
LOCUS      136901 bp  DNA  linear  PRI 30-JAN-2004
DEFINITION Homo sapiens BAC clone RP11-95P9 from 7, complete sequence.
ACCESSION AC073626
VERSION   AC073626.7  GI:12863221
KEYWORDS  HTG.
SOURCE    Homo sapiens
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
           1 (bases 1 to 136901)
           Hillier,L.W., Fulton,R.S., Fulton,L.A., Graves,T.A., Pepin,K.H.,
           Wagner-McPherson,C., Layman,D., Maas,J., Jaeger,S., Walker,R.,
           Wyllie,K., Sekhon,M., Becker,M.C., O'Laughlin,M.D., Schaller,M.E.,
           Fewell,G.A., Delehaanty,K.D., Miner,T.L., Nash,W.E., Cordes,M.,
           Du,H., Sun,H., Edwards,J., Bradshaw-Cordum,H., Ali,J., Andrews,S.,
           Isak,A., Vanbrunt,A., Nguyen,C., Du,F., Lamar,B., Courtney,L.,
           Kalicki,J., Ozerky,P., Bielicki,L., Scott,K., Holmes,A.,
           Harkins,R., Harris,A., Strong,C.M., Hou,S., Tomlinson,C.,
           Dauphin-Kohlberg,S., Kozlowicz-Reilly,A., Leonard,S., Rohlfing,T.,
           Rock,S.M., Tin-Wollam,A.M., Abbott,A., Minx,P., Maupin,R.,
           Strommatt,C., Latreille,P., Miller,N., Johnson,D., Murray,J.,
           Woessner,J.P., Wendl,M.C., Yang,S.P., Schultz,B.R., Wallis,J.W.,

```

```

Spieth,J., Bieri,T.A., Nelson,J.O., Berkowicz,N., Wohldmann,P.E.,
Cook,L.L., Hickenbotham,W.T., Eldred,J., Williams,D., Bedell,J.A.,
Mardis,E.R., Clifton,S.W., Chissoe,S.L., Marra,M.A., Raymond,C.,
Haugen,E., Gillett,W., Zhou,Y., James,R., Phelps,K., Iadamoto,S.,
Bubb,K., Simms,E., Levy,R., Clendenning,J., Kaul,R., Kent,W.J.,
Purey,T.S., Baertsch,R.A., Brent,M.R., Keibler,E., Flicek,P.,
Bork,P., Suyama,M., Bailey,J.A., Portnoy,M.E., Torrents,D.,
Chinwalla,A.T., Glsh,W.R., Eddy,S.R., McPherson,J.D., Olson,M.V.,
Eichler,E.E., Green,E.D., Waterston,R.H. and Wilson,R.K.
The DNA sequence of human chromosome 7
Nature 424 (6945), 157-164 (2003)
22737999
12853948
2 (bases 1 to 136901)
Bauer,H., Haakenson,B. and Nguyen,C.
The sequence of Homo sapiens BAC clone RP11-95P9
Unpublished (2001)
3 (bases 1 to 136901)
Waterston,R.H.
Direct Submission
Submitted (27-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 136901)
Waterston,R.H.
Direct Submission
Submitted (16-FEB-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 136901)
Waterston,R.
Direct Submission
Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
6 (bases 1 to 136901)
Waterston,R.
Direct Submission
Submitted (29-APR-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
7 (bases 1 to 136901)
Wilson,R.
Direct Submission
Submitted (30-JAN-2004) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 16, 2001 this sequence version replaced gi:11597125.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: saplens@wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0095P09
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information


```
FEATURES
source
Location/Qualifiers
1. .89728
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-134K1"
/clone_lib="RPC1-11.1"

ORIGIN

Query Match 88.9%; Score 564.6; DB 9; Length 89728;
Best Local Similarity 94.2%; Pred. No. 5.4e-167;
Matches 598; Conservative 0; Mismatches 34; Indels 3; Gaps 1;

QY 1 CCCTGTATCTTTAACTCTCTTGTAGTTGCTCTCTCCAGATCAAAATGTAATACTA 60
Db 83434 CCCTGTATCTTTAACTCTCTTGTAGTTGCTCTCTCCAGATCGAGCTGTAA---TA 83490

QY 61 CAAATTGTTCTTCAATGGAGCACAGATGGATGCCATGCTAAGATCCACCTGGACCC 120
Db 83491 TATATGGTTCTTCAATGGAGCCACAGATGCAGTCCATGACTAAGATCTACCGCGGACCC 83550

QY 121 CTGACCGGCTCTAGCCCATCTCCGATGTTAATGACATTCGAGGCACCCCTCCCGAG 180
Db 83551 CTGACCGGCTCTAGCCCATCTCCGATGTTAAGACATTCGAGGCACCCCTCCCTGAG 83610

QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
Db 83611 GAAATCTCAACTGCACAACTACTATGCCCAATTCAGCAGGAGCAGTTAGAGCGGT 83670

QY 241 CATCAGCAACCTCCCAACAGACATTTGGTTTCTGTTGAGAGGGGAGCTGAGAGAC 300
Db 83671 TGTGAGCAACCTCCCAACAGACATTTGGTTTCTGTTGAGAGGGGAGCTGAGAGAC 83730

QY 301 AGACTAGCTGGATTTCTTAGGCCACAGAGATCCCTAAGCTAGCTAGCTGGAGAGTGACT 360
Db 83731 AGACTAGCTGGATTTCTTAGGCCGGAATAAGATCCCTAAGCTAGCTGGAGAGTGACT 83790

QY 361 GCATCCACCTCTAAACATGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
Db 83791 GCATCCACCTTTAAACAGCGGGCTTGCAACTTAGCTCACACCCCAATCAGAGAGCTC 83850

QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCTTG 480
Db 83851 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCTTG 83910

QY 481 AGAGCAGCGGAGGACAGGATCGGGATATAAACCAGGATTCGAGCCGCAACGG 540
Db 83911 AGAGCAGCGGAGGACAGGATCAGGATATAAACCAGGATTCGAGCCGCAACAG 83970

QY 541 CAACCCCTTTGGTCCCTCCCTTTGTTATGGGCGCTCTGTTTCACTCTATTTCACCT 600
Db 83971 CAACCCCTTTGGTCCCTCCCTTTGTTATGGGAGCTCTGTTTCACTCTATTTCACCT 84030

QY 601 ATTAAATCTTGCACTCAAAAAAATAAAAAA 635
Db 84031 ATTAAATCTTGCACTCAAAAAAATAAAAAA 84065

RESULT 13
AC072023
LOCUS
DEFINITION Homo sapiens 3 BAC RP11-3J2 (Roswell Park Cancer Institute Human
BAC library) complete sequence.
ACCESSION AC072023
VERSION AC072023.9 GI:27764634
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 112405)
AUTHORS Muzny,D.W., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaraturunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.P.,
Bouck,J., Bowie,S., Brieva,M., Brown,B., Brown,M., Bryant,N.D.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J.J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Mezker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,
Nickerson,E., Nwokkwo,S., Ogih,M., Okwuonu,G., Oragunye,N.,
Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Saverly,G., Scherer,S.,
Scott,G., Shen,H., Shoshitari,N., Sisson,I., Sodergren,E.,
Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
Vera,V., Vallalao,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
Warren,R., Washington,C., Watlington,S., Williams,G.,
Williamson,A., Wleczkyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstein,G. and
Gibbs,R.
TITLE
JOURNAL
REFERENCE 2 (bases 1 to 112405)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 112405)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 112405)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 112405)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (30-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 6 (bases 1 to 112405)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT
On Jan 16, 2003 this sequence version replaced gi:21539113.
```

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
<http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES

```

source          Location/Qualifiers
1. .112405
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="3"
   /clone="RP11-3J2"
misc_feature    1. .2000
   /notes="overlaps bases 38072. .40071 of clone AC125608"
   /function="clone overlap"
repeat_region  complement(1..261)
repeat_region  /rpt_family="L1M4"
repeat_region  complement(431..913)
repeat_region  /rpt_family="L1ME"
repeat_region  complement(914..1221)
repeat_region  /rpt_family="AluSg"
repeat_region  complement(1222..1808)
repeat_region  /rpt_family="L1ME"
repeat_region  complement(2045..2198)
repeat_region  /rpt_family="L1ME"
repeat_region  2454..2811
repeat_region  /rpt_family="THE1C"
repeat_region  2813..3092
repeat_region  /rpt_family="MLT1I"
repeat_region  3194..3870
repeat_region  /rpt_family="L1PA3"
repeat_region  3871..3891
repeat_region  /rpt_family="AT_rich"
repeat_region  complement(4297..4671)
repeat_region  /rpt_family="MSTD"
repeat_region  4714..4913
repeat_region  /rpt_family="MLT1I"
repeat_region  5632..5670
repeat_region  /rpt_family="AT_rich"
repeat_region  complement(6337..6630)
repeat_region  /rpt_family="AluSx"

```

```

repeat_region  7063..7093
repeat_region  /rpt_family="(TTTA)n"
repeat_region  complement(7094..7374)
repeat_region  /rpt_family="AluSp"
repeat_region  complement(7386..7695)
repeat_region  /rpt_family="AluSx"
repeat_region  7910..8035
repeat_region  /rpt_family="MIR"
repeat_region  8824..10387
repeat_region  /rpt_family="L1PA16"
repeat_region  10388..10412
repeat_region  /rpt_family="AT_rich"
repeat_region  10637..10675
repeat_region  /rpt_family="AT_rich"
repeat_region  10754..10818
repeat_region  /rpt_family="L1PB1"
repeat_region  complement(10830..10905)
repeat_region  /rpt_family="L1PA16"
repeat_region  complement(10956..11184)
repeat_region  /rpt_family="L1"
repeat_region  11295..11658
repeat_region  /rpt_family="L2"
repeat_region  complement(12489..12812)
repeat_region  /rpt_family="L1PA15"
repeat_region  complement(12819..13533)
repeat_region  /rpt_family="L1"
repeat_region  13599..13619
repeat_region  /rpt_family="AT_rich"
repeat_region  14620..14933
repeat_region  /rpt_family="AluSg"
repeat_region  complement(15183..15254)
repeat_region  /rpt_family="L2"
repeat_region  complement(16463..17175)
repeat_region  /rpt_family="L1P"
repeat_region  17173..18635
repeat_region  /rpt_family="L1PA7"
repeat_region  18793..19462
repeat_region  /rpt_family="L1M4"
repeat_region  19461..20130
repeat_region  /rpt_family="L1ME"
repeat_region  20776..21073
repeat_region  /rpt_family="AluSx"

Query Match      88.2%; Score 560.2; DB 9; Length 112405;
Best Local Similarity 93.8%; Pred. No. 1.4e-165;
Matches 594; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 1 CCTGTATCTTTTAACTCTCTTGTAGTTTGTCTCTTCCAGATCAAACTGTAABACTA 60
Db 67581 CCTGTATCTTTTAACTCTCTTGTAGTTTGTCTCTTCCAGATCAAACTGTAABACTA 67640

QY 61 CAAATGTTCTTCAAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
Db 67641 CAAATGGGTCTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 67700

QY 121 CTGGACCGCTGCTAGCCCTCTCCGATGTTTAATGACATTGAAGCACCCCTCCCGAG 180
Db 67701 CTGGACTGGCTGCTAGCCCTCTCCGATGTTTAATGACATTGAAGCACCCCTCCCGAG 67760

QY 181 GAAATCTCACTGCACAAACCCCTACTATGCCCAATTCAGCGGAGGAGCTAGAGCGGT 240
Db 67761 GAAATCTCACTGCACAAACCCCTACTATGCCCAATTCAGCGGAGGAGCTAGAGCGGT 67820

QY 241 CATCAGCAACCTCCCAACAGCAGCTTGGTGTCTCTGTTGAGAGGGGAGCTAGAGAGAC 300
Db 67821 CATTGCCCAACCTCCCAACAGCAGCTTGGATCTTCTGTTGAGAGGGGAGCTAGAGAT 67880

QY 301 AGGACTAGCTGGATTTTCTTAGGCCAACGAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACT 360
Db 67881 AGGACTAGCTGGATTTTCTTAGGCCAACGAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACC 67940

QY 361 GATCCACCTCTTAAACATGGGGCTTGCAGCTTAGCTCAGACCCGAGCAATCAGAGAGCTC 420

```

Db 67941 GCATCCACCTTTAAACACACGGGGCTCGCACTTAGCTCACACCAACCAATCAGAGAGCTC 68000
QY 421 ACTAATATGCTAATTAGGCAAAATAGGAGGTAAGAAATAGCAATCATCTATTCCTG 480
Db 68001 ACTAATATGCTAATTAGGCAAAACAGGAGGTAAA-TAATAGCAATCATCTATTCCTG 68059
QY 481 AGAGCACAGCGGAGGAGCAAGGATCGGATATATAACCCAGGCGATTCGAGCGGCAACGG 540
Db 68060 AGAGCACAGCGGAGGAGGACAGGATATAAACCAGGCGATTCGAGCGGCAATGG 68119
QY 541 CAACCCCTTTGGGTCCTCCCTCCCTTGTATGGGCGCTCTGTTTCACTCTATTTCACTCT 600
Db 68120 CAACCCCTTTGGGTCCTCCCTCCCTTGTATGGGCGCTCTGTTTCACTCTATTTCACTCT 68179
QY 601 ATTAATCTTGCACTGAAAAAATAAAAAA 633
Db 68180 ATTAATCTTGCACTGAAAAAATAAAAAA 68212

RESULT 14
BS000045/c 179114 bp DNA linear PRI 12-JUN-2004
LOCUS Pan troglodytes chromosome 22 clone:RP43-179P23, map 22, complete
DEFINITION sequences.
ACCESSION BS000045 BA000046
VERSION BS000045.1 GI:37537312
KEYWORDS HTG.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE 1
1 The International Chimpanzee Chromosome 22 Consortium.
DNA sequence and comparative analysis of chimpanzee chromosome 22
Nature 429, 382-388 (2004)
2 (bases 1 to 179114)
Hattori,M., Toyoda,A., Watanabe,H., Taylor,T.D., Kuroki,Y.,
Fujiyama,A. and Sakaki,Y.
Direct Submission
Submitted (12-MAY-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
[E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170]
The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
*Chinese National Human Genome Center at Shanghai, Shanghai, China;
*GBF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
of Molecular Biotechnology, Jena, Germany; *KRIBB Genome Research
Center, Daejeon, Korea;
*Max-Planck-Institute for Molecular Genetics, Berlin, Germany;
*National Institute of Genetics, Mishima, Japan;
*National Yang Ming University Genome Research Center, Taipei,
Taiwan;
*RIKEN Genomic Sciences Center, Yokohama, Japan.
----- Genomic Center
Center: RIKEN Genomic Sciences Center
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name:The Chimpanzee Chromosome 22 Sequencing Project
Center clone name: RP43-179P23
----- Summary Statistics
Sequencing vector: pUC18, pUC13, pTZ19; 100% of reads Chemistry:
Dye-terminator Big Dye and ET; 100% of reads Assembly program:
Phrap; version 0.990329
Consensus quality: 178,810 bases at least Q40
Consensus quality: 263 bases at least Q30
Consensus quality: 35 bases at least Q20

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phased quality >=

30);
an attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by at one
plasmid
subclone or more than one M13 subclone;
and the assembly was confirmed by restriction digest.

Source information:
The RPCI-43 chimpanzee BAC library was prepared from DNA isolated
from the blood of a single male chimpanzee using published
protocols (Osoegawa,K. et al. Genomics 52:1-8). The DNA from the
chimpanzee ('Clint') was obtained from the Yerkes Primate Center in
Atlanta. The library was prepared by Baoli Zhu, Chung Li Shu,
Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library
characteristics are described at
http://www.chori.org/bacpac/mchimp43.htm.
The clone may be obtained from Pieter J. de Jong and coworkers
(http://www.chori.org/bacpac).
VECTOR: pBAC3.6
The CHORI-251 chimpanzee BAC library was prepared from DNA isolated
from the blood of a single male chimpanzee using published
protocols (Osoegawa,K. et al. Genomics 52:1-8). The DNA from the
chimpanzee ('Clint') was obtained from the Yerkes Primate Center in
Atlanta. The library was prepared by Baoli Zhu, Chung Li Shu,
Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library
characteristics are described at
http://www.chori.org/bacpac/chimpanzee251.htm.
The clone may be obtained from Pieter J. de Jong and coworkers
(http://www.chori.org/bacpac).
VECTOR: pTARBAC2.1
The PTB1 chimpanzee BAC library was prepared from DNA isolated from
cultured cells established from the blood of a single male
chimpanzee.
Clones may be obtained from Asao Fujiyama and co-workers
(http://www.gsc.riken.go.jp).
VECTOR: pKS145
The PTF22 chimpanzee Fosmid library was prepared from DNA isolated
from cultured cells established from the blood of a single male
chimpanzee.
Clones may be obtained from Asao Fujiyama and co-workers
(http://www.gsc.riken.go.jp).
VECTOR: pKS143

Sequence Quality Assessment:
This entry has been annotated with sequence
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in
10,000 bp.

Neighboring clones: RP43-014B20(left) and PTB-042H12(right).

FEATURES
Location/Qualifiers
1. .179114
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="22"
/clone="RP43-179P23"
/clone_lib="RPCI-43 chimpanzee BAC"

ORIGIN
Query Match 88.2%; Score 559.8; DB 9; Length 179114;
Best Local Similarity 94.0%; Pred. No. 2e-165;
Matches 582; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
QY 1 CCTGTATCTTTAACTCCCTTGTGTTAAAGTTGTTCTTCCAGAAATCAAACTGTAATACTA 60
Db 169185 CCTGTATCTTTAACTCCCTTGTGTTAAAGTTGTTCTTCCAGAAATCAAACTGTAATACTA 169126
QY 61 CAAATGTTCTTCAAATGGAGCACCAGATGAGTCCATGATCAAGATCCACCGTGGACCC 120
Db 169125 CAAATGTTCTTCAAATGGAGCACCAGATGAGTCCATGATCAAGATCCACCGTGGACCC 169066

```
QY 121 CTGACCGGCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGCACCCTCCCGAG 180
Db 169065 CTAGACTGGGCTGCTAGCCCATGCTCCGATTAATGACATCGAAGACACCCCTCCCGAG 169006
QY 181 GAAATCTCAACTGACAAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
Db 169005 GAAATCTCAACTGACAAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 168946
QY 241 CATCAGCAACCTCCCAACAGCAGCTTGGGTTTCTGTTTGAAGGGGGGACTGAGAGAC 300
Db 168945 CATTGGCAACCTCCCAACAGCAGCTTGGGTTTCTGTTTGAAGGGGGGACTGAGAGAC 168886
QY 301 AGGACTAGCTGGATTTCTTAGGCCAACGAAATCCCTAGCCCTAGCTAGCTGGAGAGTGA 360
Db 168885 AGGACTAGCTGGATTTCTTAGGCCAACGAAATCCCTAGCCCTAGCTAGCTGGAGAGTGA 168826
QY 361 GCATCCACCTCTAAACATGGGGCTTGCAACTAGCTCAGACCCGACCAATCAGAGAGCTC 420
Db 168825 GCATCCACCTTTAAACATGGGGCTTGCAACTAGCTCAGACCCGACCAATCAGAGAGCTC 168765
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 480
Db 168765 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 168706
QY 481 AGAGCACAGCGGAGGACAGGATCGGGATATAAACCAGGACATTCGAGCGGCAACGG 540
Db 168705 AGAGCACAGGAGGAGGACAGGATCGGGATATAAACCAGGACATTCGAGCGGCAACGG 168646
QY 541 CAACCCCTTTGGTCCCTCCCTCTGTTATGGGGCTCTGTTTCACTCTATTTCACCTCT 600
Db 168645 CAACCCCTTTGGTCCCTCCCTCTGTTATGGGGCTCTGTTTCACTCTATTTCACCTCT 168586
QY 601 ATTAATCTTGCAACTGAA 619
Db 168585 ACTAAATCTTGCAACTGCA 168567
```

RESULT 15

AP001538/c
LOCUS 174019 bp DNA linear PRI 25-MAR-2000
DEFINITION Homo sapiens genomic DNA, chromosome 21q21.1-q21.2 clone:B853K11,
1L56-APP region, complete sequence.

ACCESSION AP001538.1 GI:7328982

VERSION HTG.

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Hattori.M., Ishii.K., Toyoda.A., Taylor.T.D., Hong-Seog.P.,

Fujiyama.A., Yada.T., Totoki.Y., Watanabe.H. and Sakaki.Y.

Homo sapiens 174,019 genomic DNA of 21q21.1-q21.2

Published Only in Database (2000)

2 (bases 1 to 174019)

Hattori.M., Ishii.K., Toyoda.A., Taylor.T.D., Hong-Seog.P.,

Fujiyama.A., Yada.T., Totoki.Y., Watanabe.H. and Sakaki.Y.

Direct Submission

Submitted (23-MAR-2000) Masahira Hattori, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

Kitasato Univ., 1-15-1 Kitasato, Sagami-hara, Kanagawa 228-8555,

Japan (E-mail: hattori@gsc.riken.go.jp,

URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923,

Fax: 81-42-778-9924)

Location/Qualifiers

1. 174019

/organism="Homo sapiens"

/mol type="genomic DNA"

/db xref="taxon:9606"

/chromosome="21"

/map="21q21.1-q21.2"

/clone="B853K11"

ORIGIN

Query Match 87.7%; Score 556.6; DB 9; Length 174019;
Best Local Similarity 93.7%; Pred. No. 2.1e-164;
Matches 580; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

```
QY 1 CCTGTATCTTTAACTCTCTTGTAGTTTGTCTCTCCAGAAATCAAACTGTAATACTA 60
Db 36710 CCTGTATCTTTAACTCTCTTGTAGTTTGTCTCTCCAGAAATCAAACTGTAATACTA 36651
QY 61 CAAATTTGTTCTTCAAAATGGAGCACAGATGGAGTCCATGATGAATGATCCACCGTGGACCC 120
Db 36650 CAAATTTGTTCTTCAAAATGGAGCACAGATGGAGTCCATGATGAATGATCCACCGTGGACCC 36591
QY 121 CTGAGCGGCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCAACCCCTCCCGAG 180
Db 36590 CTAGACTGGGCTGCTAGCCCATGCTCCGATTAATGACATCGAAGGCAACCCCTCCCGAG 36531
QY 181 GAAATCTCAACTGACAAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
Db 36530 GAAATCTCAACTGACAAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 36471
QY 241 CATCAGCAACCTCCCAACAGCAGCTTGGGTTTCTGTTTGAAGGGGGGACTGAGAGAC 300
Db 36470 CATTGGCAACCTCCCAACAGCAGCTTGGGTTTCTGTTTGAAGGGGGGACTGAGAGAC 36411
QY 301 AGGACTAGCTGGATTTCTTAGGCCAACGAAATCCCTAGCCCTAGCTAGCTGGAGAGTGA 360
Db 36410 AGGACTAGCTGGATTTCTTAGGCCAACGAAATCCCTAGCCCTAGCTAGCTGGAGAGTGA 36351
QY 361 GCATCCACCTCTAAACATGGGGCTTGCAACTAGCTCAGACCCGACCAATCAGAGAGCTC 420
Db 36350 GCATCCACCTTTAAACATGGGGCTTGCAACTAGCTCAGACCCGACCAATCAGAGAGCTC 36291
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 480
Db 36290 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 36231
QY 481 AGAGCACAGCGGAGGACAGGATCGGGATATAAACCAGGACATTCGAGCGGCAACGG 540
Db 36230 AGAGCACAGCGGAGGACAGGATCGGGATATAAACCAGGACATTCGAGCGGCAACGG 36171
QY 541 CAACCCCTTTGGTCCCTCCCTCTGTTATGGGGCTCTGTTTCACTCTATTTCACCTCT 600
Db 36170 CAACCCCTTTGGTCCCTCCCTCTGTTATGGGGCTCTGTTTCACTCTATTTCACCTCT 36111
QY 601 ATTAATCTTGCAACTGAA 619
Db 36110 ACTAAATCTTGCAACTGCA 36092
```

RESULT 16

AP001674/c

LOCUS

340000 bp DNA linear PRI 21-MAY-2003

DEFINITION Homo sapiens genomic DNA, chromosome 21q, section 18/105.

ACCESSION AP001674 AL163219 BA000005

VERSION AP001674.1 GI:7768666

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1

Hattori.M., Fujiyama.A., Taylor.T.D., Watanabe.H., Yada.T.,

Park.H.S., Toyoda.A., Ishii.K., Totoki.Y., Choi.D.K., Soeda.E.,

Ohki.M., Takagi.T., Sakaki.Y., Taudien.S., Blechschmidt.K.,

Polley.A., Menzel.J., Delabar.J., Kumpf.K., Lehmann.R.,

Patterson.D., Reichwald.K., Rump.A., Schillhabel.M., Schudy.A.,

Zimmermann.W., Rosenthal.A., Kudoh.J., Shibuya.K., Kawasaki.K.,

Asakawa.S., Shintani.A., Sasaki.K., Nagamine.K., Mitsuyama.S.,

Antonarakis.S.E., Minoshima.S., Shimizu.N., Nordseik.G.,

Hornischer.K., Barand.P., Scharfe.M., Schoen.O., Desario.A.,

Reichelt.J., Kauer.G., Bloeker.H., Ramser.J., Beck.A., Klages.S.,

Hennig.S., Riesselmann.L., Dagand.E., Wehrmeyer.S., Borzym.K.,

Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Yaspo,M.L.
The DNA sequence of human chromosome 21
Nature 405 (6784), 311-319 (2000)
20289799
10830953
2 (bases 1 to 340000)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Tokoki,Y., Choi,D.K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Reichwald,K., Rump,A., Schilhabel,M., Schudy,A., Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K., Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S., Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G., Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S., Hennig,S., Rieselmann,L., Bagand,E., Wehrmaier,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Yaspo,M.L.
Direct Submission
Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)
On May 30, 2000 this sequence version replaced gi:7717271.
The chromosome 21 mapping and sequencing consortium consisting of
* RIKEN Genomic Sciences Center, Human Genome Research Group, * Sagamihara 228-8555, Japan,
* e.mail: hattori@gsc.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Molecular Biology, * Tokyo 160-8582, Japan,
* e.mail: nehimi@medb.med.keio.ac.jp
* URL: http://www.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail: info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
* Innestrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/
AL163219: Submitted (10-Apr-2000).
Location/Qualifiers
1. 340000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
1. 6149
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R749F18, 5', partial"
/clone_lib="RPCI-11 BAC library"
/notes="Accession No. AP001537"
1. 165231
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Yaspo,M.L.
The DNA sequence of human chromosome 21
Nature 405 (6784), 311-319 (2000)
20289799
10830953
2 (bases 1 to 340000)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Tokoki,Y., Choi,D.K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Reichwald,K., Rump,A., Schilhabel,M., Schudy,A., Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K., Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S., Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G., Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S., Hennig,S., Rieselmann,L., Bagand,E., Wehrmaier,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Yaspo,M.L.
Direct Submission
Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)
On May 30, 2000 this sequence version replaced gi:7717271.
The chromosome 21 mapping and sequencing consortium consisting of
* RIKEN Genomic Sciences Center, Human Genome Research Group, * Sagamihara 228-8555, Japan,
* e.mail: hattori@gsc.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Molecular Biology, * Tokyo 160-8582, Japan,
* e.mail: nehimi@medb-med.keio.ac.jp
* URL: http://www.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail: info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
* Innestrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/
AL163219: Submitted (10-Apr-2000).
Location/Qualifiers
1. 340000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
1. 6149
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R749F18, 5', partial"
/clone_lib="RPCI-11 BAC library"
/notes="Accession No. AP001537"
1. 165231
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

source
/chromosome="21"
/map="21q21.1"
/clone="R853K11, 5', partial"
/clone_lib="RPCI-11 BAC library"
/notes="Accession No. AP001538"
136796. 315056
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R845I12"
/clone_lib="RPCI-11 BAC library"
/notes="Accession No. AF245226"
156433. 277809
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="P646P19"
/clone_lib="RPCI1.3-5 PAC library"
/notes="Accession No. AL157359"
207594. 250588
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="Q3D19"
/clone_lib="LL21NCO2-Q Cosmid library"
/notes="Accession No. AL049753"
276958. 301481
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R755I19"
/clone_lib="RPCI-11 BAC library"
/notes="Accession No. AP001345"
299634. 340000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R28M9, 3', partial"
/clone_lib="RPCI-11 BAC library"
/notes="Accession No. AP000431"
311751. 340000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="C103E2331, 3', partial"
/clone_lib="ICRFc103 Cosmid library"
/notes="Accession No. AL132993"
15. 444
/note="LIMA4A"
/rpt_family="LINE/L1"
/rpt_type="DISPERSED"
585. 3531
/note="LIMA4A"
/rpt_family="LINE/L1"
/rpt_type="DISPERSED"
3770. 3876
/note="L2"
/rpt_family="LINE/L2"
/rpt_type="DISPERSED"
complement(5645. 6178)
/note="LIMA2"

repeat_region
/note="LIMA4A"
/rpt_family="LINE/L1"
/rpt_type="DISPERSED"
585. 3531
/note="LIMA4A"
/rpt_family="LINE/L1"
/rpt_type="DISPERSED"
3770. 3876
/note="L2"
/rpt_family="LINE/L2"
/rpt_type="DISPERSED"
complement(5645. 6178)
/note="LIMA2"

repeat_region
/note="LIMA4A"
/rpt_family="LINE/L1"
/rpt_type="DISPERSED"
585. 3531
/note="LIMA4A"
/rpt_family="LINE/L1"
/rpt_type="DISPERSED"
3770. 3876
/note="L2"
/rpt_family="LINE/L2"
/rpt_type="DISPERSED"
complement(5645. 6178)
/note="LIMA2"

repeat_region
/note="LIMA4A"
/rpt_family="LINE/L1"
/rpt_type="DISPERSED"
585. 3531
/note="LIMA4A"
/rpt_family="LINE/L1"
/rpt_type="DISPERSED"
3770. 3876
/note="L2"
/rpt_family="LINE/L2"
/rpt_type="DISPERSED"
complement(5645. 6178)
/note="LIMA2"

```
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
6181..6534
/notes="THE1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
8021..8370
/notes="THE1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(8373..10032)
/notes="L1MA2"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
10213..10520
/notes="AluJb"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
15810..15851
/notes="(CA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(16628..16697)
/notes="MER53"
/rpt_family="DNA"
/rpt_type=DISPERSED
16638..16715
/notes="MER53"
/rpt_family="DNA"
/rpt_type=DISPERSED
18847..19585
/notes="L2"
/rpt_family="LINE/L2"

Query Match      87.7%; Score 556.6; DB 9; Length 340000;
Best Local Similarity 93.7%; Pred No. 2.4e-164;
Matches 580; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1 CCTGTATCTTTAACTCCTTGTAAAGTTTGTCTTCCAGAAATCAAACTGTAATACTA 60
DB 27921 CCTGTATCTTTAACTCCTTGTAAAGTTTGTCTTCCAGAAATCAAACTGTAATACTA 27862
QY 61 CAAATGTTCTTCAATGGAGCACCAGATGGAGTCCATGATCAATGAATCCACCGTGAACCC 120
DB 27861 CAAATGTTCTTCAATGGAGCACCAGATGGAGTCCATGATCAATGAATCCACCGTGAACCC 27802
QY 121 CTGGACCGGCTGTAGCCCATGCTCCGATGTTAATGACATTCAGAGCACCCTCCCGAG 180
DB 27801 CTAGACTGGCTGTAGCCCATGCTCCGATGTTAATGACATTCAGAGCACCCTCCCGAG 27742
QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCCAAATTCAGCGGGAAGCAGTTAGAGCGGT 240
DB 27741 GAAATCACAACCTGCCCAACCCCTACTATGCCCCAAATTCAGAGGAGCAGTTAGAGCTGT 27682
QY 241 CATCAGCACAACCTCCCAACAGACACTTGGGTTTCTGTGAGAGGGGAGCTGAGAGAC 300
DB 27681 CATTGGCAACCTCCCAACAGACACTTGGGTTTCTGTGAGAGGGGAGCTGAGAGAC 27622
QY 301 AGGACTAGCTGATTTCTTAGGCCAACGAGAGATCCCTAAGCCTAGCTGGGAGGTTGACT 360
DB 27621 AGGACTAGCTGATTTCTTAGGCCTAAGAGATCCCTAAGCCTAGCTGGGAGGTTGACT 27562
QY 361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
DB 27561 GCATCCACTTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 27502
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAGAAATAGCAATCATCATTCCTGCTG 480
DB 27501 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAGAAATAGCAATCATCATTCCTGCTG 27442
QY 481 AGAGCAGCGGAGGAGGAGGAGGATCGGATATAAACCAGGATTCGAGCCGCGCAACGG 540
DB 27441 AGAGCAGCAGGAGGAGGAGGATCAGGATGTAAACCCAGGATTTCAAGCTCGCAACGG 27382

QY 541 CAACCCCTTTGGTCCCTCCCTTGTATGGCGCTCTGTTTCACTCTATTTCCTCT 600
DB 27381 CAACCCCTTTGGTCCCTCCCTTGTATGGCGCTCTGTTTCACTCTATTTCCTCT 27322
QY 601 ATTAAATCTTGCACACTGAA 619
DB 27321 ACTAAATCTTGCACACTGCA 27303

RESULT 17
AP001545
LOCUS 149755 bp DNA linear HTG 13-JUL-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-762G24 map.18q21, WORKING
DRAFT SEQUENCE, 13 unordered pieces.
ACCESSION AP001545
VERSION GI:9188503
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 149755)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 149,755 genomic DNA of 18q21
JOURNAL Published Only in DataBase (2000)
REFERENCE 2 (bases 1 to 149755)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT On Jul 14, 2000 this sequence version replaced gi:8117383.
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-762G24
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 145596 bases at least Q40
Consensus quality: 147102 bases at least Q30
Consensus quality: 147971 bases at least Q20
Insert size: 148555; sum-of-contigs
Quality coverage: 10.98x in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
13 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 44828 contig of 44828 bp in length
44929 82602 contig of 37674 bp in length
82703 101344 contig of 18642 bp in length
101445 116205 contig of 14761 bp in length
116306 123784 contig of 7479 bp in length
123885 130355 contig of 6471 bp in length
130456 133619 contig of 3164 bp in length
133720 137416 contig of 3697 bp in length
137517 141540 contig of 4024 bp in length
141641 144351 contig of 2711 bp in length
144452 146667 contig of 2216 bp in length
```

146768 148219 contig of 1452 bp in length
 148320 149755 contig of 1436 bp in length.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 13 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 44828: contig of 44828 bp in length
 * 44829: gap of 100 bp
 * 44929: contig of 37674 bp in length
 * 82503: gap of 100 bp
 * 82702: gap of 100 bp
 * 82703: contig of 18642 bp in length
 * 101345: gap of 100 bp
 * 101444: gap of 100 bp
 * 116205: contig of 14761 bp in length
 * 116306: gap of 100 bp
 * 123784: contig of 7479 bp in length
 * 123785: gap of 100 bp
 * 123885: contig of 6471 bp in length
 * 130456: gap of 100 bp
 * 130456: contig of 3164 bp in length
 * 133620: gap of 100 bp
 * 133720: contig of 3697 bp in length
 * 137417: gap of 100 bp
 * 137516: gap of 100 bp
 * 141540: contig of 4024 bp in length
 * 141541: gap of 100 bp
 * 141641: contig of 2711 bp in length
 * 144351: gap of 100 bp
 * 144352: contig of 2216 bp in length
 * 144522: gap of 100 bp
 * 146668: contig of 1452 bp in length
 * 146768: gap of 100 bp
 * 148220: contig of 1436 bp in length.
 * 148320: contig of 1436 bp in length.

FEATURES

source

1. 149755
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18q21"
 /clone="RP11-762G24"

1. 44828
 /note="assembly_fragment"
 44829.82602
 /note="assembly_fragment"
 82703.101344
 /note="assembly_fragment"
 101445.116205
 /note="assembly_fragment"
 116306.123784
 /note="assembly_fragment"
 123885.130355
 /note="assembly_fragment"
 130456.133619
 /note="assembly_fragment clone_end:T7 vector_side:left"
 133720.137416
 /note="assembly_fragment"
 137517.141540
 /note="assembly_fragment clone_end:SP6 vector_side:right"
 141641.144351
 /note="assembly_fragment"
 144452.146667
 /note="assembly_fragment"
 146768.148219
 /note="assembly_fragment"
 148320.149755
 /note="assembly_fragment"

ORIGIN

Query Match 86.9%; Score 551.8; DB 2; Length 149755;
 Best Local Similarity 91.8%; Pred. No. 6.8e-163;

QY	Matches	583;	Conservative	0;	Mismatches	52;	Indels	0;	Gaps	0;
1	CCCTGATCTTTAACTCTCTTGAAGTTTGTCTTCCAGAAATCAAACTCTAAACTTA	60								
67067	CCCTGATCTTTAACTCTCTTGAAGTTTGTCTTCCAGAAATCAAACTCTAAACTTA	67126								
61	CAAAATGTTTCTTAAATGGAGCAGATGGAGTCCATGACTAAGATCCACGTTGAGACC	120								
67127	CAAAATGTTTCTTAAATGGAGCAGATGGAGTCCATGACTAAGATCCACGTTGAGACC	67186								
121	CTGGACCGGCTCTAGCCCATGCTCCGATGTTAATGATTAAGGCAACCCCTCCCGAG	180								
67187	CTGGACCGGCTCTAGCCCATGCTCCGATGTTAATGATTAAGGCAACCCCTCCCGAG	67246								
181	GAATCTCAACTGCAACACCCCTACTATGCCCAATTCAGCGGGAAGCAGTTAGAGCGGT	240								
67247	GAATCTCAACTGCAACACCCCTACTATGCCCAATTCAGCGGGAAGCAGTTAGAGCGGT	67306								
241	CATCAGCCAACTCCCAACAGACACTTGGTTTTCTGTGAGAGGGGGGACTGAGAGAC	300								
67307	CATCAGCCAACTCCCAACAGACACTTGGTTTTCTGTGAGAGGGGGGACTGAGAGAC	67366								
301	AGGACTAGCTGATTTCTTAGGCAACAGAAATCCCTAAGCTAGCTGGGAAGGTGACT	360								
67367	AGGACTAGCTGATTTCTTAGGCAACAGAAATCCCTAAGCTAGCTGGGAAGGTGACT	67426								
361	GCATCCACCTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC	420								
67427	GCATCCACCTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC	67486								
421	ACTAAATGCTAATTTAGGCAAAATAGGAGTTAAGAAATAGCCATCATCTATTCGCTG	480								
67487	ACTAAATGCTAATTTAGGCAAAATAGGAGTTAAGAAATAGCCATCATCTATTCGCTG	67546								
481	AGAGCACACGGGAGGACAAAGATCGGATATAAACCCAGGCATTTCGAGCGGCAACGG	540								
67547	AGAGCACACGGGAGGACAAAGATCGGATATAAACCCAGGCATTTCGAGCGGCAACGG	67606								
541	CAACCCCTTTGGGTCCTCCCTTTGTATGGCGCTCTGTGTTTCACTCTATTTCACTCT	600								
67607	CAACCCCTTTGGGTCCTCCCTTTGTATGGCGCTCTGTGTTTCACTCTATTTCACTCT	67666								
601	ATTAAATCTTGCACTGAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA	635								
67667	TCAACTGCAAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA	67701								

RESULT 18
 AC090313/c 152980 bp DNA linear HTG 27-MAR-2003
 LOCUS Homo sapiens chromosome 18 clone RP11-762G24 map 18.
 DEFINITION
 AC090313
 AC090313.12 GI:22123575
 HTG; HTGS PHASE2; HTGS_FULLTOP; HTGS_CANCELLED.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 152980)
 Birren, B., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 18, clone RP11-762G24
 Unpublished
 2 (bases 1 to 152980)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Boguski, L., Boukhgalter, B., Brown, A.,
 Camarata, J., Campopiano, A., Cho, P., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S.,
 Dodge, S., Fero, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J.,
 Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-pierre, N.,
 Hages, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T.,
 Lechoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,
 Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,

McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Sougne, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, N., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 152980)

Barnen, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Birna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, N., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Aug 6, 2002 this sequence version replaced gi:21659527.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L12582

Center clone name: 762_G_24

* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submittor.

* This sequence will be replaced

* by the finished sequence as soon as it is available and the accession number will be preserved.

* * 1 152980: contig of 152980 bp in length.

FEATURES
source

1. 152980

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="18"

/map="18"

/clone="RP11-762G24"

/clone_lib="RPC1-11 Human Male BAC"

ORIGIN

Query Match 86.9%; Score 551.8; DB 2; Length 152980;

Best Local Similarity 91.8%; Pred. No. 6.8e-163;

Matches 583; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 1 CCTGTATCTTTAACTCCTCTTGTAAAGTTTGTCTTCCAGAAATCAAACTGTAACACTA 60
 |||||
 Db 48688 CCTGTATCTTTAACTCCTCTTGTAAAGTTTGTCTTCCAGAAATCAAAAGCTGTAACACTA 48629
 |||||
 QY 61 CAAATGTTTCTTCAAATGGAGCACCAGATGAGTCCATGACTAAGATCCACCGTGGAGCC 120
 |||||
 Db 48628 CAAATCATTTCTCAATAGGAGCCCGCCAGATGAGTCCATGACTAAGATCTTACCAGACCC 48569
 |||||
 QY 121 CTGGACCGGCTGCTAGCCCATCTCCGATGTTTAAATGACATTAAGAGCACCCCTCCCGAG 180
 |||||
 Db 48568 CTGGACCGGCTGCTAGCCCATCTCCGATGTTTAAATGACATTCGAAGGCAACCCCTCCCAAG 48509
 |||||
 QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCCAAATTCAGCGGGAAGCAGTTAGAGCGGT 240
 |||||
 Db 48508 GAAATCTCAACTGCACAAACCCCTACTATGCCCCAGTTTCAGCAGGAAGCAGTTAGGGCGGT 48449
 |||||
 QY 241 CATCAGCCAACTCTCCCAACAGCAGTCTGGTCTTCTGTTGAGAGGGGGGACTGAGAGAC 300
 |||||
 Db 48448 CATCAGCCAACTCTCCCAACAGCAGTCTGGTCTTCTGTTGAGAGGGGGTACTGAGAGAC 48389
 |||||
 QY 301 AGGACTAGCTGGATTTCTTAGGCCAACGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
 |||||
 Db 48388 AGGACTAGCTGGATTTCTTAGGGCTGACTAAGAAATCCCTAAACCTAGCTGGGAAGGTGACT 48329
 |||||
 QY 361 GCATCCACCTCTAAACATGCGGCTTGCACTTAGCTCACAACCCGACCAATCAGAGAGCTC 420
 |||||
 Db 48328 GCATTCACCTTTAAACACGAGGCTTGCACTTAGCTCACAACCTGACCAATCAGAGAGCTC 48269
 |||||
 QY 421 ACTTAAATCTTAATAGGCGAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 480
 |||||
 Db 48268 ACTTAAATCTTAATAGGCGAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 48209
 |||||
 QY 481 AGAGCACAGCGGAGGACAAAGATCGGATATAAAACCCAGGCAATTCGAGCCGCGCAACGG 540
 |||||
 Db 48208 AGAGCACAGTGGGAGGACAAAGATCAGGATATAAAACCCAGGCAATTCGAGCCGCGCAACGG 48149
 |||||
 QY 541 CAACCCCTTTGGGTCCTCCCTTGTATGGCGCTCTGTTTCTACTCTATTTCATCTCT 600
 |||||
 Db 48148 CAACCCCTTTGGGTCCTCCCTTGTATGGGAGCTCTGTTTCTACTCTATTTCATCTCT 48089
 |||||
 QY 601 ATTAATCTTGCACTGMAAAAAAAAAAAAAAAAAAAAA 635
 |||||
 Db 48088 TCAACTGCAAAATAAAAAAAAAAAAAAAAAAAAA 48054
 |||||

RESULT 19
AC093531/c
LOCUS Homo sapiens chromosome 5 clone RP11-405L7, complete sequence.
DEFINITION AC093531
ACCESSION AC093531
VERSION AC093531.2 GI:16945981
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 163803) Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 163803) DOE Joint Genome Institute.
AUTHORS Direct Submission
TITLE Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL Submitted (16-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 163803) DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS Direct Submission
TITLE Submitted (16-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL On Nov 16, 2001 this sequence version replaced gi:15383820.
COMMENT Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.3.

FEATURES

Location/Qualifiers
1. .163803
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-405L7"

ORIGIN

Query Match 86.9%; Score 551.8; DB 9; Length 163803;
Best Local Similarity 93.2%; Pred. No. 6.9e-163;
Matches 577; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
QY 1 CCTGTATCTTTAACTCCTCTGTTAAAGTTTGTCTCTCCAGAAATCAAACTGTAAACTA 60
Db 117348 CCTGTATCTTTAACTCCTCTGTTAAAGTTTGTCTCTCCAGAAATTTGAAGCTGTAAAGCTA 117289
QY 61 CAAATTTCTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGACCC 120
Db 117288 CAAATAGTTCTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGACCC 117229
QY 121 CTGACCGGCTGCTGACCCATCTCCGATCTGATGTTAAATGACATTGAAGCGCACCCCTCCCGAG 180
Db 117228 CTGACCGGCTGCTGACCTATGCTCTGATGTTAAATGACATTGAAGTCAACCCCTCCCGAG 117169
QY 181 GAAATCTCAACTGCAACAACCCCTACTATGCCCCCAATTCAGCGGGAAGCGATTAGACGGT 240
Db 117168 GAAATCTCAACTGCAACAACCCCTACTACTCCCAATTCAGTAGGAAGCGATTAGACGAGT 117109
QY 241 CATCAGCAACCTCCCAACAGACACTGGGTTTCTGTTGAGAGGGGAGCTGAGAGAC 300
Db 117108 TGTACGACCAACCTCCCAACAGACTTGGGTTTCTGTTGAGAGGGTGGACTGAGAGAC 117049
QY 301 AGGACTAGCTGGATTTCTTAGGCCAACGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
Db 117048 AGGACTAGCTGGATTTCTTAGGCTGACTAGAAATCCCTAAGCCTAGCTGGGAAGGTGACC 116989
QY 361 GCATCACCCTCTAAACATGGGCTTGCACTTAGTCTCACACCCGACCAATCAGAGAGCTC 420
Db 116988 GCATCACCCTCTTTAAACATGGGCTTGCACTTAGTCTCACACCCGACCAATCAGAGAGCTC 116929
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCTG 480
Db 116928 ACTAAATGCTAATCAGGCAAAACAGGAGTTAAGCAATAGCAATCATCTATTGCTG 116869
QY 481 AGAGCAGCGGGAGGACAAAGATCGGGATATAAACCAGGCAATTCGAGCCGCAACGG 540
Db 116868 AGAGCAGCGGGAGGACAAAGATTCGGATATAAACCAGGCAATTCAGCCGCAACAG 116809
QY 541 CAAACCCCTTTGGTCCCTCCCTTTGATGGGCTCTGTTTCACTCTATTTCATCT 600
Db 116808 CAAACCCCTTTGGTCCCTCCCTTTGATGGGAGTCTGTTTCACTCTATTTCATCT 116749
QY 601 ATTAATCTTCAACTCAA 619
Db 116748 ATTAATCTTCAACTGCA 116730

RESULT 20

AC107075/c 176425 bp DNA linear PRI 12-JUN-2002
LOCUS AC107075
DEFINITION Homo sapiens BAC clone RP11-452N17 from 2, complete sequence.
ACCESSION AC107075
VERSION AC107075.4 GI:21263352
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 176425)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
9847074
2 (bases 1 to 176425)
Swearengen-Shahid, S., Meyer, R. and Dignan, G.
The sequence of Homo sapiens BAC clone RP11-452N17
Unpublished (2001)
3 (bases 1 to 176425)
Waterston, R.H.
Direct Submission
Submitted (15-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 176425)
Waterston, R.H.
Direct Submission
Submitted (30-MAY-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 176425)
Waterston, R.
Direct Submission
Submitted (12-JUN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 30, 2002 this sequence version replaced gi:18640713.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_NH0452N17

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenos, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-624F4; the clone sequenced to the right is RP11-362J3. Actual start of this clone is at base position 1 of RP11-452N17; actual end is at base position 176425 of RP11-452N17.

Polymorphisms exist between AC110086 and AC107075. Data from AC110086, AC116626 and AC108872 was used to finish AC107075.

FEATURES

source

Location/Qualifiers
1. .176425
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"

/clone="RP11-452N17"
/clone_lib="RPC1-11"

repeat_region

2423..2523
/rpt_family="MIR"

repeat_region

3564..3598
/rpt_family="AT_rich"

repeat_region

4522..4567
/rpt_family="AT_rich"

repeat_region

4679..4756
/rpt_family="(TATA)n"

repeat_region

5700..5733
/rpt_family="(A)n"

repeat_region

5973..6006
/rpt_family="(A)n"

repeat_region

6123..6144
/rpt_family="AT_rich"

repeat_region

6803..6896
/rpt_family="L1"

repeat_region

6923..6999
/rpt_family="L1"

repeat_region

7230..7215
/rpt_family="MaLR"

repeat_region

7852..7966
/rpt_family="Tc2"

repeat_region

7966..8026
/rpt_family="Tc2"

repeat_region

8043..8172
/rpt_family="Tc2"

repeat_region

8174..8247
/rpt_family="(TATATG)n"

repeat_region

8250..8456
/rpt_family="Tc2"

repeat_region

8922..9642
/rpt_family="L1"

repeat_region

11590..11632
/rpt_family="AT_rich"

repeat_region

12173..12321
/rpt_family="MER1_type"

repeat_region

12323..12343
/rpt_family="(A)n"

repeat_region

13107..14677
/rpt_family="L1"

repeat_region

14678..14704
/rpt_family="(TTTG)n"

repeat_region

14708..14996
/rpt_family="Alu"

repeat_region

14997..15173
/rpt_family="L1"

repeat_region

15615..15720
/rpt_family="CT-rich"

repeat_region

15811..16031
/rpt_family="MaLR"

repeat_region

16093..16330
/rpt_family="L1"

repeat_region

16724..17718
/rpt_family="L1"

repeat_region

18878..19038
/rpt_family="Alu"

repeat_region

19298..19551
/rpt_family="Alu"

repeat_region

19552..19912
/rpt_family="MaLR"

repeat_region

23674..23705
/rpt_family="(CAGA)n"

repeat_region 25248..25839
/rpt_family="MaLR"
repeat_region 27499..27532
/rpt_family="AT_rich"
repeat_region 27693..27840
/rpt_family="MER1_type"
repeat_region 27902..28090
/rpt_family="MIR"
repeat_region 28091..28254
/rpt_family="MER1_type"
repeat_region 28283..28477
/rpt_family="MER1_type"
repeat_region 28510..28533
/rpt_family="AT_rich"
repeat_region 28629..28649
/rpt_family="AT_rich"
repeat_region 29707..29751
/rpt_family="(TA)n"
repeat_region 30414..30847
/rpt_family="L1"
repeat_region 31385..31942
/rpt_family="L1"
repeat_region 32421..32591
/rpt_family="L2"
repeat_region 32592..32983
/rpt_family="MaLR"
repeat_region 32984..33714
/rpt_family="L2"
repeat_region 34919..35339
/rpt_family="MaLR"
repeat_region 35752..36049
/rpt_family="Alu"
repeat_region 36274..36330
/rpt_family="Alu"
repeat_region 36767..36809
/rpt_family="AT_rich"
repeat_region 37396..37437
/rpt_family="AT_rich"
repeat_region 37816..37850
/rpt_family="A-rich"
repeat_region 38954..39375
/rpt_family="L1"
repeat_region 40130..40216
/rpt_family="L2"
repeat_region 41723..42015
/rpt_family="Alu"

Query Match 86.9%; Score 551.8; DB 9; Length 176425;
Best Local Similarity 92.8%; Pred. No. 7e-163;
Matches 593; Conservative 0; Mismatches 37; Indels 9; Gaps 1;

QY 1 CCCTGTATCTTTAACTCTTGTAAAGTTGTCTTCCAGATCAAACTGTAATACTA 60
DB 161068 CCTATATCTTTAACTCTTGTAAAGTTGTCTTCCAGATCAAACTGTAATACTA 161009
QY 61 CAAATGTTCTTCAATGGAGCACCATGAGTCCATGATCAAGATCCACCGTGGACCC 120
DB 161008 CAAATGTTCTTCAATGGAGCCCCAGATGTCAGTCCATGATCAAGATCCACCGTGGACCC 160949
QY 121 CTGGACCGGCTCTAGCCCATGCTCCGATGTTAATGACATTGAAGCACCCCTCCCGAG 180
DB 160948 CTGGACCGGCTCTAGCCCATGCTCTGATGTTAATGACATTGAAGCACCCCTCCCGAG 160889
QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAGACAGTGTAGAGCGGT 240
DB 160888 GAAATCTCAACTGCACCAACCCCTACTATGCCCGAGTTCAACAGAGACATGTAGAGCGGT 160829
QY 241 CATCAGCCAACTCTCCCAACAGACATTTGGGTTTTTCTGTTGAGAGGGGGGACTCAGAGAC 300
DB 160828 CTTGGGCCAACTCTCCCAAGACATTTGGGTTTTTCTGTTGAGAGGGGGGACTCAGAGAC 160769
QY 301 AGGACTAGCTGGATTTTCTTAGGCCAACAGAAATCCCTAGCCCTAGCTGGGAAAGGTGACT 360

Db	160768	AGGACTAGCTGGATTTCTTCCTAGGCGGACGTAAGAATCCCTTAGCTGCGGAAGTGACC	160700
Qy	361	GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCA-----	412
Db	160708	GCTTCCATCTTTAAACACGGGGCTTACAACCTTAACATCACATGACCAATCAGATAGTAA	160649
Qy	413	-GAGAGCTCACTAAATGCTTAATTAGGCAGAAAAATAGGAGGTAAAGAAATAGCCAAATCATC	471
Db	160648	GGAGAGCTCACTAAATGCTTAATTAGGCAGAAAAATAGGAGGTAAAGAAATAGCCAAATCACC	160599
Qy	472	TATTGCCTCAGAGCAGCAGCGGAGGAGCAAGATGATGGGATATAAACCCAGGCATTCGAGC	531
Db	160588	TGTTGCCTCAGAGCAGCAGCGGAGGAGCAATGATCGGATATAAACCCAGGCATTCAGC	160529
Qy	532	CGGCAACGGCAACCCCTTTGGTCCCTCCCTTTGATATGGCGCTCTGTTTTCACCTCTA	591
Db	160528	CGGCAACGGCAACCCCTTTGGTCCCTCCCTTTGATATGGAGCTCTGTTTTCACCTCTA	160469
Qy	592	TTTCACCTCTATTAATCTTGGCACTGAAAAAATAAAAAA 630	
Db	160468	CTTCACCTCTATCAATCTTGGCACTGAAAAAATAAAAAA 160430	
RESULT 21			
LOCUS	AL139090	128468 bp	DNA linear PRI 12-APR-2001
DEFINITION	Human DNA sequence from clone RP11-12B13 on chromosome 6 Contains STSs and GSSs, complete sequence.		
ACCESSION	AL139090		
VERSION	AL139090.11	GI:11228536	
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Garner, P.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: Clonequest@sanger.ac.uk		
COMMENT	On Nov 20, 2000 this sequence version replaced gi:11125403. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em', EMBL; Sw', SWISSPROT; Tr', TREMBL; Wp', WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6 IMPORTANT: This sequence is not the entire insert of clone RP11-12B13 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-12B13 is at 1 in this sequence. The true left end of clone RP11-13D18 is at 128369 in this sequence. The true right end of clone RP11-13D18 is at 128369 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-12B13 is from the library RPC1-11.1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm		

```
repeat_region 23421..23472
misc_feature /note="MIR repeat: matches 90..141 of consensus"
repeat_region 24395..24791
misc_feature /note="match: GSS: Em:AQ027280"
repeat_region 24638..24848
misc_feature /note="HAL1 repeat: matches 428..638 of consensus"
repeat_region 25008..25217
misc_feature /note="105 copies 2 mer aa 55% conserved"
repeat_region 25328..25391
misc_feature /note="32 copies 2 mer ta 70% conserved"
repeat_region 25482..26261
misc_feature /note="LIP3 repeat: matches 5..776 of consensus"
repeat_region 26257..31504
misc_feature /note="LIP3 repeat: matches 900..6146 of consensus"
repeat_region 31614..32312
misc_feature /note="LIM4 repeat: matches 3918..4608 of consensus"
repeat_region 32398..32650
misc_feature /note="LIM4 repeat: matches 4626..4909 of consensus"
repeat_region 32799..32963
misc_feature /note="FRAM repeat: matches -2..162 of consensus"
repeat_region 33151..33174
misc_feature /note="12 copies 2 mer tt 95% conserved"
repeat_region 33176..33500
misc_feature /note="LIR17 repeat: matches 1..326 of consensus"
repeat_region 33501..33791
misc_feature /note="HERV17 repeat: matches 8244..8523 of consensus"
repeat_region 33811..38137
misc_feature /note="HERV17 repeat: matches 1651..5933 of consensus"
repeat_region 38135..39957
misc_feature /note="HERV17 repeat: matches 1..1836 of consensus"
repeat_region 39958..40489
misc_feature /note="LIR17 repeat: matches 257..780 of consensus"
repeat_region 40490..41321
misc_feature /note="LIM3A repeat: matches 5261..6131 of consensus"
repeat_region 41354..41655
misc_feature /note="ALU repeat: matches 1..302 of consensus"
repeat_region 41971..42279
misc_feature /note="AluX repeat: matches 1..303 of consensus"
repeat_region 43255..43365
misc_feature /note="LIM4 repeat: matches 2169..2266 of consensus"
repeat_region 43366..43774
misc_feature /note="MSTA repeat: matches 1..426 of consensus"
repeat_region 43775..44133
misc_feature /note="LIM4 repeat: matches 2266..2691 of consensus"
repeat_region 44140..44339
misc_feature /note="LIM4 repeat: matches 6093..6289 of consensus"
repeat_region 46709..47216
misc_feature /note="LIMB8 repeat: matches 5658..6169 of consensus"
repeat_region 47217..47431
misc_feature /note="LIM4 repeat: matches 4637..4856 of consensus"
repeat_region 47487..47839
misc_feature /note="MLT2E repeat: matches 8..345 of consensus"
repeat_region 47848..48031
misc_feature /note="LIM4 repeat: matches 4451..4645 of consensus"
repeat_region 48017..49417
misc_feature /note="LIMC repeat: matches 2373..3522 of consensus"
repeat_region 49436..50712
misc_feature /note="LIMD2 repeat: matches 5056..6333 of consensus"
misc_feature complement(50811..51360)
repeat_region /note="match: GSS: Em:AQ536659"
repeat_region 52289..56182
misc_feature /note="LIP3 repeat: matches 2255..6146 of consensus"
repeat_region 56710..57178
misc_feature /note="MLT1C repeat: matches 1..463 of consensus"
repeat_region 57746..57868
misc_feature /note="MER33 repeat: matches 202..323 of consensus"
repeat_region 57869..58554
misc_feature /note="MER44C repeat: matches 1..727 of consensus"
repeat_region 58555..58752
misc_feature /note="MER33 repeat: matches 1..202 of consensus"
repeat_region 58890..59200
misc_feature /note="AluX repeat: matches 1..307 of consensus"
repeat_region 59630..60070

/note="LIMB3 repeat: matches 5737..6184 of consensus"
59956..60846
misc_feature /note="match: GSS: Em:B82923"
repeat_region 61550..61582
misc_feature /note="LIME2 repeat: matches 6133..6164 of consensus"
repeat_region 61583..62103

Query Match 86.8%; Score 551.4; DB 9; Length 128469;
Best Local Similarity 93.9%; Pred. No. 8.9e-163;
Matches 596; Conservative 0; Mismatches 36; Indels 3; Gaps 2;

Qy 1 CCCTGATCTTTAAACCTCCCTTGTAGTTTGTCTCTCCAGAAATCAAACTGTAATACTA 60
Db 33791 CCCTGATCTTTAAACCTCCCTTGTAGTTTGTCTCTCCAGAAATCAAACTGTAATACTA 33732

Qy 61 CAAATTTGTTCTTCAAAATGGAGCACAGATGAGTCCATGATCAAGATCCACCGTGGACCC 120
Db 33731 CAAATCATTTCTCAAAATGGAGCCCCAGATGAGTCCATGATCAAGATCCACCGTGGACCC 33672

Qy 121 CTGGACGGCTGCTAGCCCATGCTCCGATGTTAATGACATTTGAAGGCACCCCTCCCGAG 180
Db 33671 CT-GATCAACCTCTAGCCCATGCTCCAATGTTTAATGACATTTGAAGGCACCCCTCCCGAG 33613

Qy 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCCCAATTCAGCGGAGCAGTTAGAGCGGT 240
Db 33612 GAAATCTCAACTGCACAAACCCCTACTATGCCCCCAATTCAGCGGAGCAGTTAGAGCGGT 33553

Qy 241 CATCAGCCAACTCCCAACAGCACCTTGGTTTCTTGTGTAGAGGGGGGACTGAGAGAC 300
Db 33552 CATCAGCCAACTCCCAACAGCACCTTGGTTTCTTGTGTAGAGGGGGGACTGAGAGAC 33493

Qy 301 AGGACTAGCTGGATTTCTTAGGCCAACAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
Db 33492 AGGACTAGCTGGATTTCTTAGGCCAACAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 33433

Qy 361 GCATCCACCTTAAACATGGGCTTGCNACTTACTGCTACACCCGACCAATCAGAGACTC 420
Db 33432 GCATCCACCTTAAACATGGGCTTGCNACTTACTGCTACACCCGACCAATCAGAGACTC 33373

Qy 421 ACTAAATGCTTAATTAGGCAAAATAGGAGGTAAGAAATAGCAATCATCTATTGCTG 480
Db 33372 ACTAAGAGCTTAATTAGGCAAAATAGGAGGTAAGAAATAGCAATCATCTATTGCTG 33313

Qy 481 AGAGCACAGCGGAGGACCAAGATCGGATATAAACCAGGATTCGAGCCGCAACGG 540
Db 33312 AGAGCACAGCGGAGGACCAAGATCGGATATAAACCAGGATTCGAGCCGCAACGG 33253

Qy 541 CAACCCCTTTGGTCCCTCCCTTTGTATGGCGCTCTGTTTCTACTCTATTTCATCT 600
Db 33252 CAATCCCTTTGGTCCCTCCCTTTGTATGGCGCTCTGTTTCTACTCTATTTCATCT 33195

Qy 601 ATTAATCTTGCACTGAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 635
Db 33194 ATTAATCTTGCACTGAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT

RESULT 22
AL139038
LOCUS Human DNA sequence from clone Rp11-45618 on chromosome 13,
DEFINITION complete sequence.
ACCESSION AL139038
VERSION AL139038.18 GI:14800148
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 140756)
AUTHORS Dunn, M.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
```

```
requests: clonerequest@sanger.ac.uk
On Jul 17, 2001 this sequence version replaced gi:13751261.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-456B18 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-456B18 It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP11-456B18 is at 140756 in this
sequence. The true left end of clone RP11-78116 is at 68441 in this
sequence. The true right end of clone RP11-108H9 is at 100 in this
sequence.

FEATURES             Location/Qualifiers
     source            1..140756
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-456B18"
                     /clone_lib="RPCI-11.2"
                     1830..2326
                     /note="LOR1a repeat: matches 1..497 of consensus"
                     3932..3965
                     /note="17 copies 2 mer gt 94% conserved"
                     4632..5067
                     /note="LOR1b repeat: matches 1..461 of consensus"
                     5733..5893
                     /note="L1ME3 repeat: matches 6002..6146 of consensus"
                     6000..6083
                     /note="MER34 repeat: matches 454..539 of consensus"
                     6108..6171
                     /note="HERV23 repeat: matches 993..1056 of consensus"
                     7143..7840
                     /note="LTR8 repeat: matches 1..691 of consensus"
                     8059..8292
                     /note="LTR16A repeat: matches 203..442 of consensus"
                     8431..8686
                     /note="LTR29 repeat: matches 6..203 of consensus"
                     8725..8926
                     /note="MER4D repeat: matches 483..677 of consensus"
                     8992..9049
                     /note="MER4D repeat: matches 410..465 of consensus"
                     9082..9281
                     /note="110 copies 2 mer tt 59% conserved"
                     9398..9530
                     /note="MER41C repeat: matches 1..131 of consensus"
                     9623..9732
                     /note="LTR1 repeat: matches 675..785 of consensus"
                     9970..10333
                     /note="MLT2D repeat: matches 1..486 of consensus"

repeat_region        10400..10545
                     /note="LTR29 repeat: matches 2..151 of consensus"
repeat_region        10549..10935
                     /note="MER67D repeat: matches 12..391 of consensus"
repeat_region        11249..11332
                     /note="MIR repeat: matches 69..144 of consensus"
repeat_region        11923..12159
                     /note="MER4D repeat: matches 1..234 of consensus"
repeat_region        12164..12767
                     /note="MER4D repeat: matches 396..973 of consensus"
repeat_region        12852..13221
                     /note="THE1B repeat: matches 1..364 of consensus"
repeat_region        13222..14798
                     /note="THE1B-INTERNAL repeat: matches 1..1580 of
consensus"
repeat_region        14799..15173
                     /note="THE1B repeat: matches 1..364 of consensus"
repeat_region        15144..16346
                     /note="MER54B repeat: matches 585..789 of consensus"
repeat_region        16339..16503
                     /note="MER54B repeat: matches 112..278 of consensus"
repeat_region        16504..16571
                     /note="MER66A repeat: matches 1..69 of consensus"
repeat_region        16570..16629
                     /note="MER41B repeat: matches 283..341 of consensus"
repeat_region        16599..16902
                     /note="MER66A repeat: matches 163..478 of consensus"
repeat_region        16903..17014
                     /note="MER54B repeat: matches 3..115 of consensus"
repeat_region        17195..17491
                     /note="AluY repeat: matches 1..297 of consensus"
repeat_region        18120..19291
                     /note="LIP repeat: matches 1726..2897 of consensus"
repeat_region        19460..19584
                     /note="25 copies 5 mer atttt 62% conserved"
repeat_region        19686..19745
                     /note="12 copies 5 mer tttaa 71% conserved"
repeat_region        20453..20914
                     /note="L1MD2 repeat: matches 5745..6199 of consensus"
repeat_region        20913..21222
                     /note="L1MD2 repeat: matches 5277..5594 of consensus"
repeat_region        21308..21570
                     /note="LTR9 repeat: matches 357..625 of consensus"
repeat_region        21645..21984
                     /note="LTR19B repeat: matches 55..393 of consensus"
repeat_region        22081..22583
                     /note="L1MD2 repeat: matches 4726..5242 of consensus"
repeat_region        23498..23784
                     /note="AluJb repeat: matches 1..305 of consensus"
repeat_region        23873..24169
                     /note="AluSg repeat: matches 1..295 of consensus"
repeat_region        25044..25238
                     /note="MIR repeat: matches 31..238 of consensus"
repeat_region        25404..25696
                     /note="AluX repeat: matches 20..311 of consensus"
repeat_region        25832..26218
                     /note="MSTR repeat: matches 1..405 of consensus"
repeat_region        26221..27865
                     /note="MSTRC-internal repeat: matches 2..1651 of consensus"
repeat_region        27907..28659
                     /note="LIPB2 repeat: matches 5400..6155 of consensus"
repeat_region        28710..29043
                     /note="MSID repeat: matches 30..394 of consensus"
repeat_region        29634..29697
                     /note="32 copies 2 mer ta 79% conserved"
repeat_region        29729..29752
                     /note="12 copies 2 mer tg 95% conserved"
repeat_region        30688..30874
                     /note="MER61A repeat: matches 160..354 of consensus"
repeat_region        31648..31785
                     /note="L1MC4 repeat: matches 6409..6540 of consensus"
repeat_region        31813..32243
                     /note="L1MC4 repeat: matches 6629..7089 of consensus"
```



```

Insert size: 166466; sum-of-contigs
Quality coverage: 4.8 in Q20 bases; agarose-fp
Quality coverage: 4.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

```

ORIGIN

	Query Match	86.8%	Score 551.4	DB 2	Length 167366
Best Local Similarity	93.9%	Pred. No. 9.3e-163			
Matches 596	Conservative 0	Mismatches 36	Indels 3	Gaps 2	
Qy 1	CCCTGTATCTTTAACTCCCTTGTGTTAACTTGTCTCTTCCAGAAATCAAAACCTGTAATACTA	60			
Db 7535	CCCTGTATCTTTAACTCCCTTGTGTTAACTTGTCTCTTCCAGAAATGAAACCTGTAATACTA	7476			
Qy 61	CAAAATCTCTTCCAAATGGAGCACACAGATGGAGTCCATGACCTAAGATCCACCGCTGGACCC	120			

```
Insert size: 179011; 5.6% error; agarose-fp
Quality coverage: 12.84x in Q20 bases; sum-of-contigs Quality
Coverage: 13.23x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 183499: contig of 183499 bp in length.
FEATURES
    source
        1. 183499
            /location/Qualifiers
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="13"
            /clone="RP13-276D12"
            /clone_lib="RPC1-13.2"
            /misc_feature 1. .183499
            /note="assembly_fragment:05362"
ORIGIN
Query Match      86.8%; Score 551.4; DB 2; Length 183499;
Best Local Similarity 94.3%; Pred. No. 9,4e-163;
Matches 584; Conservative 0; Mismatches 31; Indels 4; Gaps 1;
QY 1 CCTGTATCTTTAACTCTCTTGTAGTTGTCTTCCAGATCAAAACTGTAAACTA 60
DB 12044 CCTGTATCTTTAACTCTCTTGTAGTTGTCTTCCAGATTAATAACTGTAAACTA 12103
QY 61 CAAATTTGTTCTTCAAATGGAGCACCAGATGCGAGTCCATGACTAAGATCCACCGTGGACCC 120
DB 12104 CAAATTTGTTCTTCAAATGGAGCACCAGATGCGAGTCCATGACTAAGATCCACCGGACCC 12163
QY 121 CTGGACCGGCTGCTAGCCCATGCTCCGATGTTAATGACATTTGAAGGACCCCTCCCGAG 180
DB 12164 CTGGACCGGCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGACCCCTCCCTGAG 12223
QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAAGCAGTTAGACGGT 240
DB 12224 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAAGCAGTTAGACGGT 12279
QY 241 CATCAGCAACCTCCCAACAGCACATTTGGTTTTCTTGTGAGAGGGGAGCTGAGAGAC 300
DB 12280 CATTGCCCAACTTCCCAACAGCACATTTGGTTTTCTTGTGAGAGGGGAGCTGAGAGAC 12339
QY 301 AGGACTAGCTGGATTTCTTAGGCCAACGAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
DB 12340 AGGACTAGTTGGATTTCTTAGGCCAACGAGAAATCCCTAAGCCTAGCTGGGAAGGTGACC 12399
QY 361 GCATCACTCTTAAACATGGGGCTTGCACTTAGCTTCACACCCGACCAATCAGAGAGCTC 420
DB 12400 GCATCACTCTTAAACATGGGGCTTGCACTTAGCTTCACACCCGACCAATCAGAGAGCTC 12459
QY 421 ACTAAATGCTAATAGGCAAAATAGGAGTAAAGAAATAGCAATCATCTATTGCGCTG 480
DB 12460 ACTAAATGCTAATAGGCAAAATAGGAGTAAAGAAATAGCAATCATCTATTGCGCTG 12519
QY 481 AGAGCACAGCGGAGGACCAAGGATCGGATATATAACCCAGGCAATTCGAGCCGCAACGG 540
DB 12520 AAGACACAGCGGAGGACCAAGGATCGGATATATAACCCAGGCAATTCGAGCCGCAACAG 12579
QY 541 CAACCCCTTTGGGTCCCTCCCTTTGATGGGGCTCTGTTTTCATCTTATTTCACTCT 600
DB 12580 CAACCCCTTTGGGTCCCTCCCTTTGATGGGGCTCTGTTTTCATCTTATTTCACTCT 12639
QY 601 ATTAAATCTTGCAACTGAA 619
DB 12640 GTTAAATCTTGCAACTGCA 12658
```

RESULT 25
AP002790/c

LOCUS
DEFINITION
Homo sapiens chromosome 18 clone RP11-732p12 map 18q21, WORKING
DRAFT SEQUENCE, 20 unordered pieces.

ACCESSION
AP002790.1 GI:9188576
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE

JOURNAL
Homo sapiens 164,211 genomic DNA of 18q21

AUTHORS

REFERENCE
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (11-JUL-2000) Masahita Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)

COMMENT

----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraft18

Center clone name: RP11-732p12

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 154344 bases at least Q40

Consensus quality: 159150 bases at least Q30

Consensus quality: 161200 bases at least Q20

Insert size: 162311; sum-of-contigs

Quality coverage: 5.05x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
20 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved 1

28707	28606	contig of	28606	bp in length
56074	55973	contig of	27267	bp in length
73802	73701	contig of	17628	bp in length
92614	92513	contig of	18712	bp in length
102272	102171	contig of	9558	bp in length
111354	111253	contig of	8982	bp in length
120268	120167	contig of	8814	bp in length
127738	127637	contig of	7370	bp in length
134368	134267	contig of	6530	bp in length
138847	138746	contig of	4379	bp in length
143366	143265	contig of	4419	bp in length
147744	147643	contig of	4278	bp in length
150443	150342	contig of	2599	bp in length
153103	153002	contig of	2122	bp in length
155325	155224	contig of	1243	bp in length
156668	156567	contig of	1665	bp in length
158433	158332	contig of	1711	bp in length
160244	160143	contig of	2478	bp in length
162822	162721	contig of	1390	bp in length
	164211	contig of		

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 28606: contig of 28606 bp in length
* 28607: gap of 100 bp
* 28707: contig of 27267 bp in length
* 55973: contig of 100 bp
* 56074: gap of 100 bp
* 73001: contig of 17628 bp in length
* 73002: gap of 100 bp
* 73003: contig of 18712 bp in length
* 92513: gap of 100 bp
* 92514: gap of 100 bp
* 102171: contig of 9558 bp in length
* 102172: gap of 100 bp
* 102272: contig of 8982 bp in length
* 111253: gap of 100 bp
* 111254: contig of 8814 bp in length
* 120167: gap of 100 bp
* 120168: gap of 100 bp
* 120267: contig of 7370 bp in length
* 120268: gap of 100 bp
* 127737: contig of 6530 bp in length
* 127738: gap of 100 bp
* 134267: contig of 4379 bp in length
* 134268: gap of 100 bp
* 138747: contig of 4419 bp in length
* 138748: gap of 100 bp
* 143265: contig of 4278 bp in length
* 143266: gap of 100 bp
* 147643: contig of 2599 bp in length
* 147644: gap of 100 bp
* 150342: contig of 2560 bp in length
* 150343: gap of 100 bp
* 153002: contig of 2122 bp in length
* 153003: gap of 100 bp
* 155224: contig of 1243 bp in length
* 155225: gap of 100 bp
* 156567: contig of 1665 bp in length
* 156568: gap of 100 bp
* 158332: contig of 1711 bp in length
* 158333: gap of 100 bp
* 160143: contig of 2478 bp in length
* 160144: gap of 100 bp
* 162024: contig of 1390 bp in length
* 162025: gap of 100 bp
* 162722: contig of 1390 bp in length
* 162822: gap of 100 bp

FEATURES

source

1. .164211
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21"
/clone="RP11-732P12"
1. .28606
/note="assembly_fragment"
28707. .55973
/note="assembly_fragment"
56074. .73701
/note="assembly_fragment"
73802. .92513
/note="assembly_fragment"
92614. .102171
/note="assembly_fragment"
102272. .111253
/note="assembly_fragment"
111354. .120167
/note="assembly_fragment"
120268. .127637
/note="assembly_fragment"
127738. .134267
/note="assembly_fragment"

misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

134368. .138746
/note="assembly_fragment clone_end.SP6 vector_side:left"
138847. .143265
/note="assembly_fragment"
143366. .147643
/note="assembly_fragment"
147744. .150342
/note="assembly_fragment"
150443. .153002
/note="assembly_fragment"
153103. .155224
/note="assembly_fragment"
155325. .156567
/note="assembly_fragment"
156668. .158332
/note="assembly_fragment"
158433. .160143
/note="assembly_fragment"
160244. .162721
/note="assembly_fragment"
162822. .164211
/note="assembly_fragment"

ORIGIN

Query Match 86.6%; Score 550.2; DB 2; Length 164211;
Best Local Similarity 91.7%; Pred.No. 2.2e-162; Indels 0; Gaps 0;
Matches 582; Conservative 0; Mismatches 53;
QY 1 CCTGTATCTTTAACTCTCTGTTAAGTTGTTCTCTTCCAGAACTCAAAACTGTAAACTTA 60
Db 6465 CCTGTATCTTTAACTCTCTGTTAAGTTGTTCTCTTCCAGAACTCAAAACTGTAAACTTA 6406
QY 61 CAAATTTGTTCTTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
Db 6405 CAAATCATTTCTTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGACCC 6346
QY 121 CTGGACCGCTCTGCTAGCCATGCTCCGATGTTAATGACATTGAAGCACCCTCCCGAG 180
Db 6345 CTGGACCGCTCTGCTAGCCATGCTCCGATGTTAATGACATTGAAGCACCCTCCCGAG 6286
QY 181 GAAATCTCAACTGCAACACCCCTACTATATGCCCAATTCAGCGGGAAGCAGTTAGACGGT 240
Db 6285 GAAATCTCAACTGCAACACCCCTACTATATGCCCAATTCAGCGGGAAGCAGTTAGACGGT 6226
QY 241 CATCAGCAACCTTCCCAACAGCAGCTTGGTGTCTCTGTTGAGAGGGGAGCTGAGAGAC 300
Db 6225 CATCAGCAACCTTCCCAACAGCAGCTTGGTGTCTCTGTTGAGAGGGGAGCTGAGAGAC 6166
QY 301 AGGACTAGCTGGATTTCTTAGGCCAACGAAGATCCCTAAGCCTAGCTGGGAAGGTGACT 360
Db 6165 AGGACTAGCTGGATTTCTTAGGCCAACGAAGATCCCTAAGCCTAGCTGGGAAGGTGACT 6106
QY 361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTTCACACCGCAATTCAGAGAGCTC 420
Db 6105 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTTCACACCGCAATTCAGAGAGCTC 6046
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATTCATCTATTGCTG 480
Db 6045 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATTCATCTATTGCTG 5986
QY 481 AGAGCAGCGGGAGGACAAAGATCGGGATATATAAACCAGGCAATTCGAGCCGGCAACGG 540
Db 5985 AGAGCAGCGGGAGGACAAAGATCAGGATATATAAACCAGGCAATTCGAGCCGGCAACGG 5926
QY 541 CAACCCCTTTTGGGTCCCTCCCTTGTATGGCGCTCTGTTTCTACTCTATTCTACTCT 600
Db 5925 CAACCCCTTTTGGGTCCCTCCCTTGTATGGCGCTCTGTTTCTACTCTATTCTACTCT 5866
QY 601 ATTAAATCTTGCACTGAAATATATATATATATATATATATATATATATATATATAT 635
Db 5865 TCAATGCAATATATATATATATATATATATATATATATATATATATATATATATAT 5831

```

RESULT 26
AC064801
LOCUS
DEFINITION Homo sapiens chromosome 18, clone RP11-120K19, complete sequence.
ACCESSION AC064801
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Ollivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,W.
Direct Submission
Submitted (22-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barn,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camaraata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tessfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,W.
Direct Submission
Submitted (06-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barn,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camaraata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tessfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,W.
Direct Submission
Submitted (30-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT
On Jul 30, 2002 this sequence version replaced gi:21700673.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8973
Center clone name: 120_K_19
-----
FEATURES
source 1..167843
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-120K19"
/clone_lib="RP11-120K19"
/complement(1059..1111)
/rpt_family="MIR"
/rpt_family="MIR"
/complement(1112..1424)
/rpt_family="AluX"
/complement(1425..1557)
/rpt_family="MIR"
/rpt_family="MIR"
/notes="<30 qual SNGL region"
1689..1693
/notes="<30 qual SNGL region"
/complement(1989..2161)
/rpt_family="MER97c"
2168..2463
/rpt_family="AluX"
/complement(2489..2622)
/rpt_family="FLAM_C"
/complement(3544..3967)
/rpt_family="MER97a"
4384..4633
/rpt_family="AluX"
/complement(5110..5257)
/rpt_family="MIR3"
/complement(5559..5697)
/rpt_family="MER117"
6353..6769
/rpt_family="MLT1C"
6770..6846
/rpt_family="LTR16C"
/complement(6847..6893)
/rpt_family="MIR"
/complement(7336..8363)
/rpt_family="LIP3"
8560..8867
/rpt_family="AluSc"
8962..9273
/rpt_family="AluX"
11717..11744
/rpt_family="(TAAA)n"
12041..12266
/rpt_family="MER50"

```

```
repeat_region 12267..12511
/rpt_family="AluSc"
repeat_region 12512..13031
/rpt_family="MER50"
repeat_region complement(114240..14530)
/rpt_family="AluJo"
repeat_region 14881..15288
/rpt_family="LTR16A1"
repeat_region 15290..15645
/rpt_family="THE1C"
repeat_region complement(15654..15794)
/rpt_family="MER5A"
repeat_region 15935..16489
/rpt_family="MLT2A2"
repeat_region complement(17190..17503)
/rpt_family="AluXb"
repeat_region complement(17524..17861)
/rpt_family="LIME2"
repeat_region complement(17998..18430)
/rpt_family="MER65A"
repeat_region 19847..19874
/rpt_family="AT rich"
repeat_region complement(19903..20222)
/rpt_family="AluJo"
repeat_region complement(21102..21388)
/rpt_family="AluJb"
repeat_region 21410..21692
/rpt_family="L2"
repeat_region 21693..21985
/rpt_family="AluXb"
repeat_region 21986..22149
/rpt_family="L2"
repeat_region 22201..22391
/rpt_family="LTR67"
repeat_region 22657..22934
/rpt_family="AluJb"
repeat_region 22940..23062
/rpt_family="LIME3"
repeat_region 24066..24354
/rpt_family="LIME4A"
repeat_region 25031..25348
/rpt_family="AluXb"
repeat_region 25583..26009
/rpt_family="MLT1C"
repeat_region 28144..28456
/rpt_family="AluY"
repeat_region complement(28729..28847)
/rpt_family="LIME/D"
repeat_region 29217..29290
/rpt_family="L2"
repeat_region 29341..29442
/rpt_family="MLT1I"
repeat_region 29557..29579
/rpt_family="(GAAGA)n"
repeat_region complement(29806..30079)
/rpt_family="AluJb"

Query Match 86.68; Score 550.2; DB 9; Length 167843;
Best Local Similarity 91.78; Pred. No. 2.2e-162;
Matches 582; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 1 CCCTGTATCTTTAACTCCCTTGTAGTTTGTCTTCCAGATCAAACTGTAAACTA 60
DB 50488 CCCTGTATCTTTAACTCCCTTGTAGTTTGTCTTCCAGATCAAACTGTAAACTA 50547

QY 61 CAAATGTTCTTCAATGGAGCACCAGATGAGTCCATGATCCACCGTGGACCC 120
DB 50548 CAAATGTTCTTCAATGGAGCACCAGATGAGTCCATGATCCACCGTGGACCC 50607

QY 121 CTGACCGGCTGCTAGCCCATGCTCGCATGTTAATGACATTTGAAGCACCCTCCCGAG 180
DB 50608 CTGACCGGCTGCTAGCCCATGCTCGCATGTTAATGACATTTGAAGCACCCTCCCGAG 50667
```

```
QY 181 GAAATCTCAACTGCACACCCCTACTATGCCCCCAATTCAGCGGAGCAGTTAGACGGT 240
DB 50668 GAAATCTCAACTGCACACCCCTACTATGCCCCCAATTCAGCGGAGCAGTTAGACGGT 50727
QY 241 CATCAGCCAACTCCCAACAGACACTTGGGTTTTCTGTGTGAGAGGGGAGCTAGAGAC 300
DB 50728 CATCAGCCAACTCCCAACAGACACTTGGGTTTTCTGTGTGAGAGGGGAGCTAGAGAC 50787
QY 301 AGGACTAGCTGGATTTCTTAGGCCAACGAAGAAATCCCTAAGCCTAGCTCGGAAGGTGACT 360
DB 50788 AGGACTAGCTGGATTTCTTAGGCCAACGAAGAAATCCCTAAGCCTAGCTCGGAAGGTGACT 50847
QY 361 GCATCCACCTTAACATGCGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGTCT 420
DB 50848 GCATCCACCTTTAAACACGCGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGTCT 50907
QY 421 ACTAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAAATCATCTATTCCTG 480
DB 50908 ACTAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAAATCATCTATTCCTG 50967
QY 481 AGAGCACAGCGGAGGACAAAGATCGGATATATAACCCAGGCAATTCGAGCCGCGCAACGG 540
DB 50968 AGAGCACAGTGGGAGGACAAAGATCAGGATATAAACCCAGGCAATTCGAGCCGCGCAACGG 51027
QY 541 CAACCCCTTTGGTCCCTCCCTTGTATGGCGCTCTGTTTCACTCTATTTCACTCT 600
DB 51028 CAACCCCTTTGGTCCCTCCCTTGTATGGCGCTCTGTTTCACTCTATTTCACTCT 51087
QY 601 ATTAAATCTTGCAACTGAAAAAATAAAAAAATAAAAAAATAAAAAA 635
DB 51088 TCNACTGCAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 51122
```

RESULT 27

```
CNS01DSH 180523 bp DNA linear PRI 20-AUG-2001
LOCUS Human chromosome 14 DNA sequence BAC R-30513 of library RPCI-11
DEFINITION from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION AL121784.5 GI:15282089
VERSION HTG; HTGS ACTIVEFIN.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 180523)
AUTHORS Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
Gyapay,G., Saurin,W. and Weissenbach,J.
TITLE Sequencing of the human chromosome 14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 180523)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2001) Genoscope - Centre National de Sequencage :
AUTHORS BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
On Aug 23, 2001 this sequence version replaced gi:12001714.
----- Genom Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr
-----
The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : C-2547L24 (AC=AL135818)
Downstream BAC (overlapping the SP6 end) : R-580E16 -----
Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 7.78x in Q20 bases; sum-of-contigs
-----
```

Overall quality chart :

Range : bases

0 : bases

1 - 9 : bases

10 - 19 : bases

20 - 29 : bases

30 - 39 : bases

40 - 49 : bases

50 - 59 : bases

60 - 69 : bases

70 - 79 : bases

80 - 89 : bases

90 - 99 : bases

Percentage of bases with a quality value >= 40 : 99 %.

FEATURES

source

Location/Qualifiers

1. .180523

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="14"

/clone_lib="R-30513"

/clone_lib="RPCI-11"

1561. .1825

/notes="matching EMBL:R74544"

RHdb:RH53560

dbSTS:STS42868

Identified using the e-PCR software (G. Schuler) "

69085. .69312

/notes="matching EMBL:G03465"

RHdb:RH53501

RHdb:RH3654

dbSTS:STS1070

Identified using the e-PCR software (G. Schuler) "

71190. .71397

/notes="matching EMBL:AA010373"

RHdb:RH75482

dbSTS:STS2570

Identified using the e-PCR software (G. Schuler) "

ORIGIN

Query Match

Best Local Similarity 86.58; Score 549; DB 9; Length 180523;

Matches 590; Conservative 0; Mismatches 35; Indels 10; Gaps 1;

QY 1 CCCTGTATCTTTAACTCTCTGTTAGTTGTTCTCTTCCAGAAATCAAACTGTAAACTA 60

DB 537 CCCTTTATCTTTAACTCTCTGTTAGTTGTTCTCTTCCAGAAATCAAACTGTAAACTA 596

QY 61 CAAATGTTCTTCAATGAGACCCAGATGGAGTCCATGACTAAGATCCACCGTGACCC 120

DB 597 CAAATGTTCTTCAATGAGACCCAGATGGAGTCCATGACTAAGATCCACCGTGACCC 656

QY 121 CTGACGGCGCTGACCCATCTCGATGTTAATGACATTTGAAGCACCCTCCCGAG 180

DB 657 CTGACGGCGCTGACCCATCTCGATGTTAATGACATTTGAAGCACCCTCCCGAG 716

QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCCAAATTCAGCGGAAGCAGTTAGACGGT 240

DB 717 GAAATCTCAACTGCACAAACCCCTACTATGCCCCAAATTCAGCGGAAGCAGTTAGACGAGT 776

QY 241 CATCAGCAACCTCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGAGACTGAGAGAC 300

DB 777 TGTGAGCAACCTCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGAGACTGAGAGAC 836

QY 301 AGCACTAGCTGGATTTCTAGGCGCAGAGATCCCTAGCTAGCTAGCTGGAGAGTACT 360

DB 837 AGCACTAGCTGGATTTCTAGGCGCAGAGATCCCTAGCTAGCTAGCTGGAGAGTACT 896

QY 361 GCATCACCTCTAAACATAGGGGCTTGCACTTAGCTCAGACCCGACCAATCAGAGAGCTC 420

DB 897 ACATCCACCTTTAAACAGAGGGTTGCACTTAGCTCAGACCCGACCAATCAGAGAGCTC 956

QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCCAAATCATCTATTGCTGT 480

DB 957 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCCAAATCATCTATTGCTGT 1016

QY 481 AGAGCACAGCGGAGGACAAAGATCGGATATATAAACCAGGCAATTCGAGCCGCAACGG 540

DB 1017 AGAGCACAGCGGAGGACAAATCATCGGATATATAAACCAGGCAATTCGAGCCGCAATGG 1076

QY 541 CAACCCCTTTGGGTCCTCCCTTTGATGGGGCTCTCTTTTCACTCTATTTCACTCT 600

DB 1077 CTACCTCTTTGGGTCCTCCCTTTGATGGGGCTCTCTTTTCACTCTATTTCACTCT 1126

QY 601 ATTAATCTTGCACCTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 635

DB 1127 ATTAATCTTGCACCTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1161

RESULT 28

AC007374

LOCUS

DEFINITION

AC007374

AC007374

AC007374.6

GI:13450000

HTG.

Homo sapiens (human)

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 190565)

Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,

Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,

Pate, D. and Hood, L.

Sequencing of human chromosome 15 D15S146-D15S117 region

Unpublished

2 (bases 1 to 190565)

Rowen, L., Madan, A., Qin, S., Abbasi, N., Dors, M., Dickhoff, R.,

James, R., Koretz, C., Lasky, S., Madan, A., Prescott, S., Ratcliffe, A.,

Shaffer, T. and Hood, L.

Direct Submission

Submitted (25-APR-1999) Multimegabase Sequencing Center, University

of Washington, PO BOX 357730, Seattle, WA 98195, USA

3 (bases 1 to 190565)

Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,

Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,

Pate, D. and Hood, L.

Direct Submission

Submitted (27-MAR-2001) Multimegabase Sequencing Center, Institute

for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA

98105, USA

On Mar 27, 2001 this sequence version replaced gi:8247780.

Genome Center

Center: Multimegabase Sequencing Center

Center code: UMSC

Web site: http://chroma.mbt.washington.edu/msg_www

Contact: leerowen@systemsbiology.org

Summary Statistics

Sequencing vector: pUC18; 108752

Chemistry: Dye-terminator Big Dye; 90% of reads

Chemistry: Dye-primer Big Dye; 10% of reads

Assembly program: Phrap; version 0.990399

Location/Qualifiers

1. .190565

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="14"

/map="14q31"

/clone="RP11-325L17"

/clone_lib="RPCI human BAC library 11"

985. .1035

unsure

```
unsure /notes="low quality data"
3865..3870
unsure /notes="low quality data"
14875..14880
unsure /notes="low quality data"
17060..17095
unsure /notes="low quality data"
17488..17535
unsure /notes="low quality data"
19433
unsure /notes="low quality data"
46474..46580
unsure /notes="low quality data"
46695..46700
unsure /notes="low quality data"
55163..55242
unsure /notes="low quality data"
110860..110875
unsure /notes="low quality data"
113613..113617
unsure /notes="low quality data"
116230..116782
misc_feature /notes="Single clone coverage. PCR and transposons tried."
unsure 117677..117680
unsure 117730..117748
unsure 118505..118736
unsure 144847..144850
unsure 152067..152071
unsure 158953..158990
unsure 177622..177705
unsure /notes="low quality data"

ORIGIN
Query Match 86.5%; Score 549; DB 9; Length 190565;
Best Local Similarity 92.94; Pred. No. 5.5e-162; Indels 10; Gaps 1;
Matches 590; Conservative 0; Mismatches 35;

QY 1 CCCTGTATCTTTAACTCCCTTGTTAGTTGTCTCTTCCAGAAATCAAACTGTAACAACTA 60
DB 23017 CCCTTATCTTTAACTCCCTTGTTAGTTGTCTCTTCCAGAAATCAAACTGTAACAACTA 23076

QY 61 CAAATGTTCTTCAATGAGACACAGATGAGTCCATGACTAAGATCCACCGTGACCC 120
DB 23077 CAAATAGTTCTTCAATGAGACACAGATGAGTCCATGACTAAGATCTACCATGACCC 23136

QY 121 CTGACCGGCTCTAGCCATCTCCGATGTTAATGACATTCAGAGCACCCCTCCCGAG 180
DB 23137 CTGACCGGCTCTAGCCATCTCCGATGTTAATGACATTCAGAGCACCCCTCCCGAG 23196

QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAGACAGTTAGAGCGGT 240
DB 23197 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAGACAGTTAGAGCAGT 23256

QY 241 CATCAGCCAACTCCCAACAGACACTTGGGTTTCTCTGTTGAGAGGGGGAGTGAAGAC 300
DB 23257 TGTGAGCCAACTCCCAACAGACACTTGGGTTTCTCTGTTGAGAGGGGGAGTGAAGAC 23316

QY 301 AGGACTAGCTGGATTTCTTAGGCCAACGAAGATCCCTAAGCTAGCTGGGAAGGTGACT 360
DB 23317 AGGACTAGCTGGATTTCTTAGGCCAACGAAGATCCCTAAGCTAGCTGGGAAGGTGACC 23376

QY 361 GCATCCACCTCTAAACATGGGGCTTGCAACTTACTAGCTCACACCCGACCAATCAGAGACTC 420
DB 23377 ACATCCACCTTTAAACACAGGGCTTGCAACTTACTAGCTCACACCCGACCAATCAGAGACTC 23436

QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAGAAATAGCAATCATCTATTGCTG 480
|||||
```

```
Db 23437 ACTAAATGCTAATTAGGCAAAACAGGAGGTAAGAAATAGCAATCATCTATCCCTG 23496
QY 481 AGAGCACAGCGGAGGACAGAGATCGGATATAAAACCCAGGCAATTCGAGCGGCAACGG 540
|||||
Db 23497 AGAGCACAGCAGGAGGACAAATGATCGGATATAAAACCCAGGCAATTCGAGCGGCAATGG 23556
|||||
QY 541 CAACCCCTTTGGTCCCTCCCTTGTATGGCGCTCTGTTTCTACTCTATTTTCACTCT 600
|||||
Db 23557 CTACCCCTTTGGTCCCTCCCTTGTATGGGAGCTCAGT-----TTTCACTCT 23606
|||||
QY 601 ATTAATCTTGCACACTGAAAAAAGAAAAA 635
|||||
Db 23607 ATTAATCTTGCACACTGAAAAAAGAAAAA 23641
|||||
```

```
RESULT 29
CNS01DVH/c 203777 bp DNA linear PRI 04-MAY-2001
LOCUS Human chromosome 14 DNA sequence BAC C-2547124 of library Caltech-D
DEFINITION from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION AL135818
VERSION AL135818.3 GI:8217905
KEYWORDS HTG. Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 203777)
AUTHORS Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
Levy,M., Eckenberg,R., Bruls,T., deSerardinis,V., Cruaud,C.,
Gyapay,G., Saurin,W. and Weissbach,J.
TITLE Sequencing of the human chromosome 14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 203777)
GENOSCOPE Genoscope.
DIRECT SUBMISSION 2 (bases 1 to 203777)
TITLE Genoscope.
JOURNAL Submitted (04-MAY-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT On Jun 3, 2000 this sequence version replaced gi:7406569.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr
-----
The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-305I3
Downstream BAC (overlapping the SP6 end) : R-895M11 (AC=AL133153)
----- Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 7.44x in Q20 bases; sum-of-contigs
-----
```

```
Overall quality chart :
Range : bases
0 : 2
1 - 9 : 29
10 - 19 : 188
20 - 29 : 490
30 - 39 : 1165
40 - 49 : 5179
50 - 59 : 5946
60 - 69 : 9708
70 - 79 : 23956
80 - 89 : 61022
90 - 99 : 96092
-----
Percentage of bases with a quality value >= 40 : 99 %.
FEATURES
Location/Qualifiers
1..203777
/organism="Homo sapiens"
```

Db	9913	ACTAAATGCTAATTAGGCAAAACAGGAGGTAAAGAAATAGCAATCATCTATCGCGCTG	9854
QY	481	AGAGCACAGCGGAGGACCAAGGATCGGATATAAACCCAGGATTCGAGCCGCAACGG	540
Db	9853	AGAGCACAGCAGGAGGACAAATGATCGGATATAAACCCAGGATTCGAGCCAGCAATGG	9794
QY	541	CAACCCCTTTGGGTCCCTCCCTTTGTATGGCGCTCTGTTTCACTCTATTTCACTCT	600
Db	9793	CTACCCCTCTTTGGGTCCCTCCCTTTGTATGGAGCTCAGT-----TTTCACTCT	9744
QY	601	ATTAAATCTTGCNACTGAAAAAATAAAAAA	635
Db	9743	ATTAAATCTTGCNACTGCAAAAACAAACAA	9709
RESULT 30			
AC109992		112544 bp	DNA
LOCUS			
DEFINITION			
AC109992			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
1 (bases 1 to 112544)			
Muzny, D.M., Adams, C.,			
Alsbrooks, S.L., Amarante, H.C.,			
Barbata, J., Benton, J., Binage, K.,			
Bouck, J., Bowie, S., Brieva, M.,			
Buhay, C., Burch, P., Burkett, C.,			
Carton, T.F., Carter, M., Cavazos, S.R.,			
Chen, G., Chen, K., Chen, Z., Chowdhry, I.,			
Cleveland, C.D., Cox, C., Coyle, M.D.,			
Davila, M.L., Davis, C., Davy-Carroll, L.,			
Delaney, K.R., Delgado, O., Denn, A.L.,			
Douchwaite, K.J., Draper, H., Dugan-Rocha, S.,			
Earnhart, C., Edgar, D., Edwards, C.C.,			
Falls, T., Ferraguto, D., Flagg, N., Ford, J.,			
Gabisi, A., Gao, J., Garcia, A., Garner, T.,			
Correll, J.H., Guevara, W., Gunaratne, P.,			
Harris, C., Harris, K., Hart, M., Havlak, P.,			
Hernandez, J., Hernandez, O., Hodgson, A.,			
Hollings, B., Homs, F., Howard, S., Huber, J.,			
Jackson, L.B., Jacobson, B., Jia, Y., Johnson, R.,			
Joudah, S., Karlsson, E., Kelly, S., Khan, U.,			
Kovar, C., Kratovic, J., Kureshi, A., Landry, N.,			
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C.,			
Louis, H., Lozano, R.J., Lu, X., Lucier, A.,			
Ma, J., Maheshwari, M., Mapua, P., Martin, R.,			
Martinez, E., Massey, B., Mawhinney, E.,			
Mei, G., Metzker, M., Miner, G., Miner, Z.,			
Moore, S., Morgan, M., Moorish, T., Morris, S.,			
Nelson, D., Newton, J., Newton, N., Nguyen, A.,			
Nickerson, E., Nwokenkwo, S., Oguh, M.,			
Oviedo, R., Pace, A., Payton, B., Peery, J.,			
Pickens, R., Primus, E., Pu, L.L., Quiles, M.,			
Rojas, A., Rojubo, I., Roife, M., Ruiz, S.,			
Scott, G., Shen, H., Shoohtari, N., Sison, I.,			
Sonaik, T., Sparks, A., Stanley, H., Stone, H.,			
Tabor, P., Tamera, A., Tamera, K., Tang, H.,			
Taylor, T., Telford, B., Thomas, N., Thomas, S.,			
Ver, V., Villalon, D., Vinson, R., Wang, Q.,			
Warren, R., Washington, C., Watlington, S.,			
Williamson, A., Wlezyk, R., Wooden, S.,			
Wu, Y.F., Zhou, J., Zorrilla, S., Zorrilla, S.,			
Gibbs, R.			
Direct Submission			
Unpublished			
2 (bases 1 to 112544)			
Worley, K.C.			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			

TITLE
JOURNAL
Direct Submission
Submitted (09-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 112544)
Worley,K.C.
REFERENCE
AUTHORS
TITLE
JOURNAL
Direct Submission
Submitted (18-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 112544)
Worley,K.C.
REFERENCE
AUTHORS
TITLE
JOURNAL
Direct Submission
Submitted (26-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 112544)
Worley,K.C.
REFERENCE
AUTHORS
TITLE
JOURNAL
Direct Submission
Submitted (25-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 26, 2002 this sequence version replaced gi:20976483.
COMMENT
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.

Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
<http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht>
ml.

FEATURES

source
1. .112544
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clones="RP11-383G6"
complement(46..1498)
/rpt_family="L1M4"
complement(1499..1804)
/rpt_family="AluSc"
complement(1805..2516)
/rpt_family="L1M4"

repeat_region
repeat_region
repeat_region

repeat_region complement(2556..2621)
/rpt_family="AluJ/FLAM"
repeat_region complement(2622..2786)
/rpt_family="L1M4"
repeat_region complement(2792..2833)
/rpt_family="L1M4"
repeat_region complement(2834..3134)
/rpt_family="AluY"
repeat_region complement(3135..3392)
/rpt_family="L1M4"
repeat_region 5391..5415
/rpt_family="AT_rich"
repeat_region complement(5509..5805)
/rpt_family="AluY"
repeat_region 6389..6551
/rpt_family="MIR"
repeat_region 6981..7109
/rpt_family="AluSg/x"
repeat_region 7173..7203
/rpt_family="TTTA)n"
repeat_region complement(7206..7491)
/rpt_family="AluSx"
repeat_region complement(7505..7618)
/rpt_family="L1MCS"
repeat_region 7761..8048
/rpt_family="AluSx"
repeat_region 10539..10837
/rpt_family="AluSg"
repeat_region 11017..11061
/rpt_family="MADE1"
repeat_region 11176..11213
/rpt_family="AT_rich"
repeat_region complement(11215..11374)
/rpt_family="UI"
repeat_region complement(11387..11681)
/rpt_family="AluSx"
repeat_region 12173..12203
/rpt_family="AT_rich"
repeat_region 12313..12350
/rpt_family="A-rich"
repeat_region 13076..13361
/rpt_family="AluSx"
repeat_region 14236..14529
/rpt_family="L1M4"
repeat_region complement(15115..15406)
/rpt_family="AluSx"
repeat_region 15454..15878
/rpt_family="L1MCS"
repeat_region 16207..16227
/rpt_family="AT_rich"
repeat_region 16946..17002
/rpt_family="AT_rich"
repeat_region 18059..18361
/rpt_family="AluJo"
repeat_region 19122..19425
/rpt_family="AluSx"
repeat_region complement(19530..20155)
/rpt_family="L2"
repeat_region complement(20156..20458)
/rpt_family="AluSg1"
repeat_region complement(20459..20559)
/rpt_family="L2"
repeat_region 20687..20790
/rpt_family="MLT1J2"
repeat_region 21217..21344
/rpt_family="L1M4"
repeat_region complement(21391..21514)
/rpt_family="MER5A"
22051..22099
/rpt_family="L2"
22410..22545

Query Match

86.3%; Score 547.8; DB 9; Length 112544;

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2005, 11:11:14 ; Search time 291.418 Seconds
(without alignments)
11438.492 Million cell updates/sec

Title: US-09-319-156b-6
Perfect score: 635
Sequence: 1 cccgtatctttaacctctc.....tgaaaaaaaaaaaaaaaaaaa 635

Scoring table: IDENTITY NUC
Gapop 10.0 , Capext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	635	100.0	635	2	Aav43215 Multiple
2	635	100.0	635	2	Aax29702 Clone Cl6
3	635	100.0	635	12	ADG14843
4	626	98.6	635	9	Adg14843 MSRV asso
5	617.4	97.2	2030	3	Adb84397 MSRV-1 as
6	546.4	86.0	2074	6	Aaa63826 Nucleotid
7	545	85.8	1329	2	Adad1225 Human EMB
8	545	85.8	1329	2	Aav43219 Multiple
9	545	85.8	1329	2	Aax29704 Clone SM6
10	540.2	85.1	2046	8	Adg14849
11	535	84.3	1329	9	ACC46747
12	533.8	84.1	2946	2	Adb84403 MSRV-1 as
13	533.8	84.1	2946	3	Aax77526 Human sec
14	533.8	84.1	2946	10	Aaz59468 Human sec
15	532.8	83.9	326014	6	Adc38776 Human cdn
16	532.8	83.9	326014	12	Abk9296 Human gen
17	532.2	83.8	2781	5	ADQ94981
18	525.8	82.8	1136	2	Aaf55630 Nucleotid
19	525.8	82.8	1136	3	Aax25660 Human end
20	522.6	82.3	2782	2	Aaa59210 3' pol ge
21	522.6	82.3	2782	3	Aax25661 Human end
					Aaa59211 5' non co

22	522.6	82.3	2782	5	AAH20069	Aah20069	HERV-W en
23	517.4	81.5	2930	6	AAD24195	Ad24195	Human syn
24	516.2	81.3	56093	6	ABL61744	Ab161744	Colon ade
25	511.4	80.5	7582	2	AAZ25665	Aax25665	Complete
26	511.4	80.5	7582	3	AAZ59215	Aaa59215	Human end
27	510.6	80.4	46340	3	ABN97978	Abn97978	Human ret
28	505.8	79.7	10499	3	ABN97929	Abn97929	Human ret
29	499.6	78.7	1894	4	AAI14608	Aai14608	Probe #45
30	499.6	78.7	1894	4	ABA56337	Ab56337	Human foe
31	499.6	78.7	1894	4	AAI35980	Aai35980	Probe #46
32	499.6	78.7	1894	4	ABA45822	Ab45822	Human bre
33	499.6	78.7	1894	4	ABA25978	Ab25978	Probe #44
34	499.6	78.7	1894	4	AAK30018	Aak30018	Human bon
35	499.6	78.7	1894	4	AAK04516	Aak04516	Human bra
36	499.6	78.7	1894	4	ABS29670	Ab29670	Human liv
37	499.6	78.7	1894	5	AAI04422	Aai04422	Probe #44
38	499.6	78.7	1894	6	ABS04589	Ab04589	Human gen
39	495.6	78.0	3372	2	AAZ25663	Aax25663	Human end
40	495.6	78.0	3372	3	AAZ59213	Aaa59213	Partial p
41	493.2	77.7	2052	4	ABA08902	Ab08902	Human rec
42	482.6	76.0	22436	9	ADA02882	Ada02882	Human PAP
43	482.6	76.0	22436	10	ADB72620	Adb72620	Human PAP
44	482.6	76.0	22436	10	ADC85361	Adc85361	Mouse Pap
45	482.6	76.0	22436	12	ADM74477	Adm74477	Human car

ALIGNMENTS

RESULT 1
AAV43215
ID AAV43215 standard; cDNA; 635 BP.
XX
AC AAV43215;
XX
DT 29-DEC-1998 (first entry)
XX
DE Multiple sclerosis associated retrovirus fragment 4.
XX
Multiple sclerosis associated retrovirus; MSRV; MS; gag gene;
KW env gene; rheumatoid arthritis-associated virus; ss.
XX
Multiple sclerosis associated retrovirus.
XX
Key Location/Qualifiers
FH CDS 1..234
FT /*tag= a
FT /product= "Encodes protein AAV71067"
XX
PN W09823755-A1.
XX
PD 04-JUN-1998.
XX
PF 26-NOV-1997; 97WO-1B001482.
XX
PR 26-NOV-1996; 96US-00756429.
XX
PA (INMR) BIO MERIEUX.
XX
PI Perron H, Beseme F, Bedin F, Paranhos-Baccala G;
PI Komurian-Pradel F, Jollivet-Reynaud C, Mandrand B;
XX
WPI; 1998-322732/28.
P-PSDB; AAV71067.
XX
PT New nucleic acid from retroviruses - useful for diagnosis, prevention and treatment of, e.g. multiple sclerosis.
XX
PS Disclosure; Page 183; 286pp; English.
XX
CC The present sequence represents a multiple sclerosis (MS) associated retrovirus (MSRV) genomic fragment used in the method of the invention.
CC The invention provides complete or partial genomic sequences of the MSRV-

CC 1 pol gene, gag gene and env gene, and polypeptides encoded by these
CC genes. The invention also provides antibodies raised against the
CC polypeptides. The genomic sequences, polypeptides and antibodies are also
CC claimed useful for diagnosing infection by MS and rheumatoid arthritis-
CC associated viruses, and also for prevention and treatment of infection
CC with these viruses
XX
SQ Sequence 635 BP; 188 A; 170 C; 136 G; 141 T; 0 U; 0 Other;

Query Match 100.0%; Score 635; DB 2; Length 635;
Best Local Similarity 100.0%; Pred. No. 9.1e-187; Indels 0; Gaps 0;
Matches 635; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCTGTATCTTTAACTCTCTTTGTTAGTTTGTCTCTTCCAGAACTCAAACTGTAAACTA 60
Db 1 CCTGTATCTTTAACTCTCTTTGTTAGTTTGTCTCTTCCAGAACTCAAACTGTAAACTA 60
QY 61 CAAATTTCTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGAACCC 120
Db 61 CAAATTTCTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGAACCC 120
QY 121 CTGACCGGCTCTGACCCATCTCCGATGTTAATGACATTAAGGACACCCCTCCCGAG 180
Db 121 CTGACCGGCTCTGACCCATCTCCGATGTTAATGACATTAAGGACACCCCTCCCGAG 180
QY 181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGACGGT 240
Db 181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGACGGT 240
QY 241 CATCAGCAACCTCCCAACAGACACTTGGTTTTCTGTGAGAGGGGGAGTGAAGAC 300
Db 241 CATCAGCAACCTCCCAACAGACACTTGGTTTTCTGTGAGAGGGGGAGTGAAGAC 300
QY 301 AGGACTAGCTGGATTTCTTAGGCCAACAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
Db 301 AGGACTAGCTGGATTTCTTAGGCCAACAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
QY 361 GCATCCACCTCTAAACATGGGGCTTGCACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
Db 361 GCATCCACCTCTAAACATGGGGCTTGCACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
QY 421 ACTAAATGCTAATAGGCAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCTG 480
Db 421 ACTAAATGCTAATAGGCAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCTG 480
QY 481 AGAGCAGCGGGAGGACAAAGATCGGGATATAAACCCAGGCATTGAGCCGGCAACGG 540
Db 481 AGAGCAGCGGGAGGACAAAGATCGGGATATAAACCCAGGCATTGAGCCGGCAACGG 540
QY 541 CAACCCCTTTGGGTCCCTCCCTTTGTATGGCGCTCTGTTTCACTCTATTTCACCTCT 600
Db 541 CAACCCCTTTGGGTCCCTCCCTTTGTATGGCGCTCTGTTTCACTCTATTTCACCTCT 600

RESULT 2
AA29702
ID AAX29702 standard; DNA; 635 BP.
XX
AC AAX29702;
XX
DT 17-OCT-2003 (revised)
DT 27-AUG-2003 (revised)
DT 08-JUN-1999 (first entry)
XX
DE Clone CL6-3' from MSRV-1.
XX Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
XX rheumatoid polyarthritis; ss.
XX

OS Viruses.
XX FR276588-A1.
XX 08-JAN-1999.
XX 07-JUL-1997; 97FR-00088816.
XX 07-JUL-1997; 97FR-00088816.
XX (INMR) BIO MERIEUX.
XX WPI; 1999-098275/09.
XX P-PSDB; AAW99552.
XX Nucleic acid sequences of retrovirus called MSRV-1 - associated with
XX multiple sclerosis or rheumatoid polyarthritis.
XX Claim 1; Page 36-37; 83pp; French.
XX This sequence represents clone CL6-3' from a novel multiple sclerosis
XX related virus type 1 (MSRV1). The sequence can be used in diagnostic,
XX prophylactic or therapeutic compositions to inhibit expression of a
XX multiple sclerosis related virus and/or virus associated with rheumatoid
XX polyarthritis. (Updated on 27-AUG-2003 to correct OS field.) (Updated on
XX 17-OCT-2003 to standardise OS field)
XX Sequence 635 BP; 188 A; 170 C; 136 G; 141 T; 0 U; 0 Other;

Query Match 100.0%; Score 635; DB 2; Length 635;
Best Local Similarity 100.0%; Pred. No. 9.1e-187; Indels 0; Gaps 0;
Matches 635; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCTGTATCTTTAACTCTCTTTGTTAGTTTGTCTCTTCCAGAACTCAAACTGTAAACTA 60
Db 1 CCTGTATCTTTAACTCTCTTTGTTAGTTTGTCTCTTCCAGAACTCAAACTGTAAACTA 60
QY 61 CAAATTTCTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGAACCC 120
Db 61 CAAATTTCTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGAACCC 120
QY 121 CTGACCGGCTCTGACCCATCTCCGATGTTAATGACATTAAGGACACCCCTCCCGAG 180
Db 121 CTGACCGGCTCTGACCCATCTCCGATGTTAATGACATTAAGGACACCCCTCCCGAG 180
QY 181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGACGGT 240
Db 181 GAAATCTCAACTGCACAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGACGGT 240
QY 241 CATCAGCAACCTCCCAACAGACACTTGGTTTTCTGTGAGAGGGGGAGTGAAGAC 300
Db 241 CATCAGCAACCTCCCAACAGACACTTGGTTTTCTGTGAGAGGGGGAGTGAAGAC 300
QY 301 AGGACTAGCTGGATTTCTTAGGCCAACAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
Db 301 AGGACTAGCTGGATTTCTTAGGCCAACAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
QY 361 GCATCCACCTCTAAACATGGGGCTTGCACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
Db 361 GCATCCACCTCTAAACATGGGGCTTGCACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
QY 421 ACTAAATGCTAATAGGCAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCTG 480
Db 421 ACTAAATGCTAATAGGCAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCTG 480
QY 481 AGAGCAGCGGGAGGACAAAGATCGGGATATAAACCCAGGCATTGAGCCGGCAACGG 540
Db 481 AGAGCAGCGGGAGGACAAAGATCGGGATATAAACCCAGGCATTGAGCCGGCAACGG 540
QY 541 CAAACCCCTTTGGGTCCCTCCCTTTGTATGGCGCTCTGTTTCACTCTATTTCACCTCT 600
Db 541 CAAACCCCTTTGGGTCCCTCCCTTTGTATGGCGCTCTGTTTCACTCTATTTCACCTCT 600

```

QY 601 ATTAATCTTGAACCTGAAAAAAGAAAAAAGAAAAA 635
DB 601 ATTAATCTTGAACCTGAAAAAAGAAAAAAGAAAAA 635

RESULT 3
ADG14843
ID ADG14843 standard; cdna; 635 BP.
XX
AC ADG14843;
XX
XX 26-FEB-2004 (first entry)
XX
DE MSRV associated cdna #7.
XX
XX ss; pol gene; retrovirus; multiple sclerosis; rheumatoid arthritis.
XX
XX Multiple sclerosis associated retrovirus.
XX
XX US2003198647-A1.
XX
XX 23-OCT-2003.
XX
XX 03-APR-2002; 2002US-00114104.
XX
XX 26-NOV-1996; 96US-00756429.
XX
XX 26-NOV-1997; 97US-00979847.
XX
XX (INMR ) BIO MERIEUX.
XX
XX Perron H, Beseme F, Bedin F, Paranhos-Baccala G;
XX Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B, Garson JA, Tuke PW;
XX WPI; 2004-032461/03.
XX
XX New isolated nucleic acid and their fragments having the pol gene of a
XX retrovirus, useful for diagnosing, preventing and/or treating multiple
XX sclerosis and/or rheumatoid arthritis.
XX
XX Disclosure; SEQ ID NO 102; 193pp; English.
XX
XX The invention relates to an isolated nucleic acid which comprises the pol
XX gene of a retrovirus associated with multiple sclerosis or rheumatoid
XX arthritis. The methods and compositions of the present invention are
XX useful for diagnosing, preventing and/or treating multiple sclerosis
XX and/or rheumatoid arthritis. The present sequence is used in the
XX exemplification of the invention.
XX
XX Sequence 635 BP; 188 A; 170 C; 136 G; 141 T; 0 U; 0 Other;

Query Match 100.0%; Score 635; DB 12; Length 635;
Best Local Similarity 100.0%; Pred. No. 9.1e-187;
Matches 635; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTGTATCTTTAACTCTTGTAAAGTTTCTCTTCCAGAAATCAAACTGTAATACTA 60
DB 1 CCTGTATCTTTAACTCTTGTAAAGTTTCTCTTCCAGAAATCAAACTGTAATACTA 60
QY 61 CAAATTTCTTCAAAATGGAGCACCAGATGAGTCCATGACTAAGATCCACCGTGGACCC 120
DB 61 CAAATTTCTTCAAAATGGAGCACCAGATGAGTCCATGACTAAGATCCACCGTGGACCC 120
QY 121 CTGGACCGGCTGCTAGCCCATGCTCCGATGTTAATGACATTTGAAGGACCCCTCCCGAG 180
DB 121 CTGGACCGGCTGCTAGCCCATGCTCCGATGTTAATGACATTTGAAGGACCCCTCCCGAG 180
QY 181 GAAATCTCAACTGACAAACCCCTACTATGCCCAATTCAGCGGAAGCACTTAGAGCGGT 240
DB 181 GAAATCTCAACTGACAAACCCCTACTATGCCCAATTCAGCGGAAGCACTTAGAGCGGT 240
QY 241 CATTAGCCCACTCCCAACAGCACTTGGGTTTCTGTTGAGGGGGGACTGAGAGAC 300
DB 241 CATCAGCCCACTCCCAACAGCACTTGGGTTTCTGTTGAGGGGGGACTGAGAGAC 300

301 AGGACTAGCTGGATTTCTTAGGCCAACGAAGATCCCTAAGCCCTAGCTGGGAAGTCACT 360
301 AGGACTAGCTGGATTTCTTAGGCCAACGAAGATCCCTAAGCCCTAGCTGGGAAGTCACT 360
361 GCATCCACCTCTAAACATGAGGGCTTTGCAACTTACCTCACAACCCGACCAATCAGAGAGCTC 420
361 GCATCCACCTCTAAACATGAGGGCTTTGCAACTTACCTCACAACCCGACCAATCAGAGAGCTC 420
421 ACTAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCTG 480
421 ACTAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCTG 480
481 AGAGCAGCGGAGGAGGACCAAGGATCGGATATTAACCCAGGCAATTCGAGCCGCAACGG 540
481 AGAGCAGCGGAGGAGGACCAAGGATCGGATATTAACCCAGGCAATTCGAGCCGCAACGG 540
541 CAACCCCTTTGGTCCCTCCCTTTGTTATGGGCGCTCTGTTTTCACCTCTATTTCACCTCT 600
541 CAACCCCTTTGGTCCCTCCCTTTGTTATGGGCGCTCTGTTTTCACCTCTATTTCACCTCT 600
601 ATTAATCTTGCACACTGAAAAAAGAAAAAAGAAAAA 635
601 ATTAATCTTGCACACTGAAAAAAGAAAAAAGAAAAA 635

RESULT 4
ADB84397
ID ADB84397 standard; DNA; 635 BP.
XX
XX ADB84397;
XX
XX 04-DEC-2003 (first entry)
XX
XX MSRV-1 associated DNA sequence #17.
XX
XX ds; multiple sclerosis; rheumatoid arthritis; gag; pol;
XX reverse transcriptase; ribonuclease H.
XX
XX Unidentified;.
XX
XX OS
XX
XX US2003039664-A1.
XX
XX 27-FEB-2003.
XX
XX 26-NOV-1997; 97US-00979847.
XX
XX 26-NOV-1996; 96US-00756429.
XX
XX (PERR/) PERRON H.
XX (BESE/) BESEME F.
XX (BEDI/) BEDIN F.
XX (PARA/) PARANHOS-BACCALA G.
XX (KOMU/) KOMURIAN-PRADEL F.
XX (JOLI/) JOLIVET-REYNAUD C.
XX (MAND/) MANDRAND B.
XX (GARS/) GARSON J A.
XX (TUKP/) TUKE P W.
XX
XX Perron H, Beseme F, Bedin F, Paranhos-Baccala G;
XX Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B, Garson JA, Tuke PW;
XX WPI; 2003-512253/48.
XX
XX New isolated or purified nucleic acid associated with multiple sclerosis
XX and/or rheumatoid arthritis, useful for detecting a virus associated with
XX multiple sclerosis or rheumatoid arthritis in a biological sample.
XX
XX Claim 31; Page 76-77; 193pp; English.
XX
XX The invention relates to an isolated or purified nucleic acid from a
XX virus associated with multiple sclerosis and/or rheumatoid arthritis,
XX multiple sclerosis-associated virus (MSRV)-1. The nucleic acids comprise

```

CC pol, gag or reverse transcriptase genes (or their fragments) encoding the
CC proteins or defined peptides (including immunodominant peptides,
CC antigenic peptides or conserved motifs). Also included are a process for
CC detecting a virus associated with multiple sclerosis or rheumatoid
CC arthritis in a biological sample, a nucleic acid probe for the detection
CC of a virus associated with multiple sclerosis or rheumatoid arthritis, a
CC primer for the amplification by polymerisation of a nucleic acid of a
CC viral material associated with multiple sclerosis or rheumatoid
CC arthritis, a polypeptide exhibiting an inhibitory activity on the
CC proteolytic, reverse transcriptase or ribonuclease H activity from MSRV,
CC and an antibody directed against the MSRV-1 virus obtained by
CC immunological reacting a human or animal body or cells with an
CC immunogenic agent consisting of the antigenic polypeptide defined above.
CC The nucleic acids are useful for detecting a biological sample a virus
CC associated with multiple sclerosis or rheumatoid arthritis, or for
CC detecting in a biological sample, the presence of or exposure to a virus
CC associated with multiple sclerosis or rheumatoid arthritis. The present
CC sequence is a claimed MSRV-associated sequence whose identity cannot be
CC accurately determined. Note: The SEQ ID numbers for the sequences as
CC displayed in the main body of the patent do not match the SEQ ID numbers
CC in the sequence listing. Consequently those sequences mentioned in the
CC claims may not be the sequences the authors intended to claim.

XX SQ Sequence 635 BP; 188 A; 167 C; 133 G; 138 T; 0 U; 9 Other;
Query Match 98.6%; Score 626; DB 9; Length 635;
Best Local Similarity 98.6%; Pred. No. 5.7e-184;
Matches 626; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 CCTGTATCTTTAACTCCCTTGTGTTAGTTTGTCTTCCAGATCAAACTGTAAACTA 60
Db 1 CCTGTATCTTTAACTCCCTTGTGTTAGTTTGTCTTCCAGATCAAACTGTAAACTA 60
Qy 61 CAATGTTCTTCAATGGAGCACCAGATGAGTCCATGATCAAGTCCACCGTGACCC 120
Db 61 CAATGTTCTTCAATGGAGCACCAGATGAGTCCATGATCAAGTCCACCGTGACCC 120
Qy 121 CTGACCGCGCTGCTAGCCCATGCTCCGATGTTAATGACATTTGAAGGACCCCTCCCGAG 180
Db 121 CTGACCGCGCTGCTAGCCCATGCTCCGATGTTAATGACATTTGAAGGACCCCTCCCGAG 180
Qy 181 GAAATCTCAACTGACAAACCCCTACTATGCCCAATTCAGCGGAGCAGTTAGCGGT 240
Db 181 GAAATCTCAACTGACAAACCCCTACTATGCCCAATTCAGCGGAGCAGTTAGCGGT 240
Qy 241 CATCAGCAACCTCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGAGCTGAGAG 300
Db 241 CATCAGCAACCTCCCAACAGCACTTGGGTTTTCCTGTTGAGAGGGGGAGCTGAGAG 300
Qy 301 AGGACTAGCTGGATTCTTAGGCCCAAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACT 360
Db 301 AGGACTAGCTGGATTCTTAGGCCCAAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACT 360
Qy 361 GCATCCACCTTAAACATGGGGCTTGCATTTAGCTCACAACCGGACCAATCAGAGACTC 420
Db 361 GCATCCACCTTAAACATGGGGCTTGCATTTAGCTCACAACCGGACCAATCAGAGACTC 420
Qy 421 ACTAAATGCTTAATAGGCAAAATAGGAGTGAAGAAATAGCAATCATCTATTGCTG 480
Db 421 ACTAAATGCTTAATAGGCAAAATAGGAGTGAAGAAATAGCAATCATCTATTGCTG 480
Qy 481 AGAGCACAGGGGAGGACAGGATCGGGATATATAACCCAGGATTCGAGCCGCGCAACGG 540
Db 481 AGAGCACAGGGGAGGACAGGATCGGGATATATAACCCAGGATTCGAGCCGCGCAACGN 540
Qy 541 CAACCCCTTTGGTCCCTCCCTTTGATGGGCGCTCTGTTTCACTCTATTTCATCT 600
Db 541 CAACCCCTTTGGTCCCTCCCTTTGATGGGCGCTCTGTTTCACTCTATTTCATCT 600
Qy 601 ATTAAATCTTGCACTGAAAAAAGAAAAAAGAAAAA 635
Db 601 ATTAAATCTTGCACTGAAAAAAGAAAAAAGAAAAA 635

RESULT 5
AAA63826
ID AAA63826 standard; DNA; 2030 BP.
XX
AC AAA63826;
XX
DT 06-AUG-2003 (revised)
DT 04-DEC-2000 (first entry)
XX
DE Nucleotide sequence of the MSRV-1 3' env and LTR regions.
XX
KW MSRV-1; pol region; long terminal repeat; LTR; RUS region; retrovirus;
KW ss.
XX
OS Multiple sclerosis associated retrovirus.
XX
FH Key Location/Qualifiers
FT CDS 1..1629
FT /tag= a
FT /note= "Contains one termination codon"
FT sig_peptide 1..81
FT /tag= b
FT CAAT_signal 1800..1807
FT /tag= c
FT CAAT_signal 1858..1864
FT /tag= d
FT TATA_signal 1906..1911
FT /tag= e
FT polyA_signal 1996..2002
FT /tag= f
XX
PN WO200047745-A1.
XX
PD 17-AUG-2000.
XX
PF 15-FEB-2000; 2000WO-IB000159.
XX
PR 15-FEB-1999; 99BP-00420041.
XX
PA (INMR) BIO MERIEUX.
XX
PI Paranhos-Baccala G, Perron H, Komurian-Pradel F;
XX
DR WPI; 2000-506097/45.
XX
PT P-PSDB; AAB08195.
XX
PT Nucleotide fragment of LTR-RUS region from Multiple Sclerosis retrovirus
XX (MSRV) used to detect the presence of MSRV-1 retrovirus in a biological
XX sample.
XX
PS Disclosure; Fig 2; 23pp; English.
XX
CC The present sequence represents the nucleotide sequence corresponding to
CC the 3' env region and long terminal repeat sequences from clone Cl6 of
CC Multiple Sclerosis retrovirus (MSRV-1). The specification describes a
CC long terminal repeat (LTR)-RUS region which encodes the expression of a
CC MSRV-1 protein. This is unusual for LTRs, in particular in the RUS
CC region. The sequence includes CAAT and TATA signals which are present in
CC the U3 and R regions and are not directed towards the MSRV-1 retrovirus
CC the features table. Probes and antibodies to the MSRV-1 retrovirus
CC protein and encoding polynucleotide sequences are used to detect the
CC presence of MSRV-1 retrovirus in a biological sample. (Updated on 06-AUG-
CC 2003 to correct OS field.)
XX
SQ Sequence 2030 BP; 574 A; 559 C; 387 G; 510 T; 0 U; 0 Other;
Query Match 97.2%; Score 617.4; DB 3; Length 2030;
Best Local Similarity 98.3%; Pred. No. 4.8e-181;
Matches 624; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
Qy 1 CCTGTATCTTTAACTCCCTTGTGTTAGTTTGTCTTCCAGATCAAACTGTAAACTA 60

Db 1396 CCTGTATCTTTAACTCCTTGTGTTAGTTTGTCTCTCCAGAATTGAGCTGTAAGCTA 1455
QY 61 CAAATGTTCTTCAATGGAGCACCAGATGAGTCCATGACTAAGATCCACCGTGGACCC 120
Db 1456 CAGATGGTCTTACAAATGGAAACCCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 1515
QY 121 CTGACCGGCTGCTAGCCCTGCTCGATGTTAATGACATTTGAAGCACCCTCCCGAG 180
Db 1516 CTGACCGGCTGCTAGCCCTGCTCGATGTTAATGACATTTGAAGCACCCTCCCGAG 1575
QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
Db 1576 GAAATCTCAACTGCACAAACCCCTACTATGCCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 1635
QY 241 CATCAGCAACCTCCCAACAGACATTTGGTTTCTGTTGAGAGGGGGACTGAGAGAC 300
Db 1636 CATCAGCAACCTCCCAACAGACATTTGGTTTCTGTTGAGAGGGGGACTGAGAGAC 1695
QY 301 AGGACTAGCTGGATTTCTTAGGCCAAGCAAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
Db 1696 AGGACTAGCTGGATTTCTTAGGCCAAGCAAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 1755
QY 361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
Db 1756 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 1815
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTCCCTG 480
Db 1816 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTCCCTG 1875
QY 481 AGAGCAGCGGAGGAGCAAGATCGGGATATAAACCAGGAGGATTCGAGCCGGCAACGG 540
Db 1876 AGAGCAGCGGAGGAGCAAGATCGGGATATAAACCAGGAGGATTCGAGCCGGCAACGG 1935
QY 541 CAACCCCTTTGGTCCCTCCCTTCTGTTATGGGGCTCTGTTTCACTCTATTTCCTCT 600
Db 1936 CAACCCCTTTGGTCCCTCCCTTCTGTTATGGGGCTCTGTTTCACTCTATTTCCTCT 1995
QY 601 ATTAATCTTGCAACTGAAAAAAMAAAAA 635
Db 1996 ATTAATCTTGCAACTGAAAAAAMAAAAA 2030
RESULT 6
AAD41225
ID AAD41225 standard; cdna; 2074 BP.
XX
AC AAD41225;
XX
XX 30-OCT-2002 (first entry)
XX Human EMBRY-2 CDNA.
XX
KW Human; embryogenesis associated protein; AIDS; reproductive disorder;
KW infertility; endometriosis; endometrial tumour; inflammatory disorder;
KW autoimmune disorder; acquired immune deficiency syndrome; transgenic;
KW ovarian tumour; contact dermatitis; placenta disorder; preclampsia;
KW EMBRY-2; allergy; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 64..1692
FT /*tag= a
FT /product= "EMBRY-2 protein"
FT sig_peptide 64..123
FT /*tag= b
FT mat_peptide 124..1689
FT /*tag= c
FT /product= "Mature EMBRY-2 protein"
XX
PN W020248362-A2.
XX

PD 20-JUN-2002.
XX
PF 14-NOV-2001; 2001WO-US043956.
XX
PR 15-NOV-2000; 2000US-0249407P.
XX (INCY-) INCYTE GENOMICS INC.
XX Ramkumar J, Arvizu C;
PI WPI; 2002-537629/57.
DR P-PSDB; AAE25054.
XX
XX New polypeptides of human embryogenesis associated proteins for screening
PT modulators useful for treating or preventing disorders e.g.
PT endometriosis, infertility, allergy, preclampsia.
XX
PS Claim 59; Page 96-97; 97pp; English.
XX
XX The invention relates to human embryogenesis associated proteins (EMBRY)
CC and nucleic acid molecules encoding such proteins. EMBRY sequences are
CC useful for screening modulators useful for treating or preventing
CC disorders associated with abnormal expression of EMBRY. The disorders
CC treated include reproductive disorders such as infertility.
CC endometriosis, endometrial or ovarian tumour; autoimmune/inflammatory
CC disorder such as acquired immune deficiency syndrome (AIDS), allergies,
CC contact dermatitis; disorders of the placenta such as preclampsia,
CC abruptio placentae etc. Sequences of the invention are also useful for
CC analysing a proteome of a tissue or a cell type. EMBRY proteins are
CC useful as immunogens for preparing antibodies. Polynucleotides of the
CC invention are useful for creating knockin humanised animals or transgenic
CC animals to model human diseases. They are also used in gene therapy. The
CC present sequence is human EMBRY-2 CDNA
XX
SQ Sequence 2074 BP; 583 A; 567 C; 392 G; 531 T; 0 U; 1 Other;
Query Match 86.0%; Score 546.4; DB 6; Length 2074;
Best Local Similarity 93.8%; Pred. No. 5.9e-159;
Matches 579; Conservative 0; Mismatches 37; Indels 1; Gaps 1;
QY 1 CCTGTATCTTTAACTCCTTGTGTTAGTTTGTCTCTCCAGAATCAAACTGTAAGCTA 60
Db 1459 CCTGTATCTTTAACTCCTTGTGTTAGTTTGTCTCTCCAGAATCAAACTGTAAGCTA 1518
QY 61 CAAATGTTCTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
Db 1519 CAAATGTTCTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGACCC 1578
QY 121 CTGACCGGCTGCTAGCCCTGCTCGATGTTAATGACATTTGAAGCACCCTCCCGAG 180
Db 1579 CTGACCGGCTGCTAGCCCTGCTCGATGTTAATGACATTTGAAGCACCCTCCCGAAG 1638
QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
Db 1639 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 1698
QY 241 CATCAGCAACCTCCCAACAGACATTTGGTTTCTGTTGAGAGGGGGACTGAGAGAC 300
Db 1699 CGTCAGTCAACCTCCCAACAGACATTTGGTTTCTGTTGAGAGGGGGACTGAGAGAC 1758
QY 301 AGGACTAGCTGGATTTCTTAGGCCAAGCAAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
Db 1759 AGGACTAGCTGGATTTCTTAGGCCAAGCAAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 1818
QY 361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
Db 1819 GCGTCCACCTTTAAACACCGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 1878
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTTCCTG 480
Db 1879 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTTCCTG 1938
QY 481 AGAGCAGCGGAGGAGCAAGATCGGGATATAAACCAGGAGGATTCGAGCCGGCAACGG 540

Db 1939 AGAGCAGCTGGGAGGAGGAGGATTCGAATATAAACCCAGGCAATTCGAGCCAGC-ANAG 1997
QY 541 CAACCCCTTTGGGTCCTCCCTTGTATGGGGCTCTGTTTCACTTACTCT 600
Db 1998 CAACCGCTTTGGGTCCTCCCTTGTATGGGAGCTCTGTTTCACTTACTCT 2057
QY 601 ATTAAATCTGCAACTG 617
Db 2058 ATTAAATCTTGCAACTG 2074

RESULT 7
AAV43219
ID AAV43219 standard; cDNA; 1329 BP.
XX AC
XX AAV43219;
XX 29-DEC-1998 (first entry)
XX Multiple sclerosis associated retrovirus fragment 6.
XX Multiple sclerosis associated retrovirus; MSRV; MS; pol gene; gag gene;
KW env gene; rheumatoid arthritis-associated virus; ss.
XX Multiple sclerosis associated retrovirus.

Key Location/Qualifiers
CDS 2..490
FT /*tag= a
FT /product= "Encodes protein AAW71069"
FT /transl_except= (pos:77-79, appears to code for a stop
FT codon)
FT /transl_except= (pos:125-127, appears to code for a stop
FT codon)
FT /transl_except= (pos:137-139, appears to code for a stop
FT codon)

XX W09823755-A1.
XX 04-JUN-1998.
XX 26-NOV-1997; 97WO-IB001482.
XX 26-NOV-1996; 96US-00756429.
XX (INMR) BIO MERIEUX.
XX Perron H, Beseme F, Bedin F, Paranhos-Baccala G;
XX Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B;
XX WPI; 1998-322732/28.
XX P-PSDB; AAW71069.
XX New nucleic acid from retroviruses - useful for diagnosis, prevention and
XX treatment of, e.g. multiple sclerosis.
XX Disclosure; Page 187-188; 286pp; English.

XX The present sequence represents a multiple sclerosis (MS) associated
XX retrovirus (MSRV) genomic fragment used in the method of the invention.
XX The invention provides complete or partial genomic sequences of the MSRV-
XX 1 pol gene, gag gene and env gene, and polypeptides encoded by these
XX genes. The invention also provides antibodies raised against the
XX polypeptides. The genomic sequences, polypeptides and antibodies are also
XX claimed useful for diagnosing infection by MS and rheumatoid arthritis-
XX associated viruses, and also for prevention and treatment of infection
XX with these viruses
XX Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 0 U; 3 Other;
Query Match 85.8%; Score 545; DB 2; Length 1329;
Best Local Similarity 92.4%; Pred. No. 1.3e-150;

Matches 572; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 1 CCTGTATCTTTAACTCTCTTGTAAAGTTTGTCTTCCAGAATCAAACTGTAAAACTA 60
Db 257 CCTGTATCTTCAACTTCTTGTAAAGTTTGTCTTCCAGAATGAAGCTGTAAAGCTA 316
QY 61 CAAATTTGTTCTTCAATGGAGCACAGATGGAGTCCATGACTAAGATCACCCTGAGACC 120
Db 317 CAAATAGTTCTTCAAAATGGAACCCAGATGCAATGCAATGCAATGCAATGCAATG 376
QY 121 CTGGACCGGCTCTAGCCCATGCTCCGATGTTAATGACATTTGAAGGACCCCTCCCGAG 180
Db 377 CTGGACCGGCTCTAGACTATGCTCTGATGTTAATGACATTTGAAGTACCCCTCCCGAG 436
QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGGAAGCAGTTAGAGCGT 240
Db 437 GAAATCTCAACTGCACAAACCCCTACTACACTCCAATTCAGTAGGAAGCAGTTAGAGCAGT 496
QY 241 CATCAGCCAACTCTCCCAACAGCACATTTGGGTTTTCTGTTGAGAGGGGGGACTGAGAGAC 300
Db 497 TGTGAGCCAACTCTCCCAACAGCACATTTGGGTTTTCTGTTGAGAGGGGGGACTGAGAGAC 556
QY 301 AGGACTAGCTGATTTCCCTAGGCCCAACGAAGATCCCTAAGCCTAGCTGGGAAGGTGACT 360
Db 557 AGGACTAGCTGATTTCCCTAGGCCCAACGAAGATCCCTAAGCCTAGCTGGGAAGGTGACT 616
QY 361 GCATCCACCTCTAAACATGCGGCTTTGCAACTTTAGCTCACACCCGACCAATCAGAGAGCTC 420
Db 617 GCATCCATCTTTAAACATGCGGCTTTGCAACTTTAGCTCACACCCGACCAATCAGAGAGCTC 676
QY 421 ACTAAATCTTAATTAGGCAAAAATAGGAGGTAAGAATAGCAATAGCAATCATCTATTGCGCTG 480
Db 677 ACTAAATCTTAATCAGGCAAAAACAGGAGGTAAGAATAGCAATAGCAATCATCTATTGCGCTG 736
QY 481 AGAGCACACGCGGAGGACAAAGATCGGATATAAACCCAGGCATTCGAGCCGCGCAACGG 540
Db 737 AGAGCACACGCGGAGGACAAAGATCGGATATAAACCCAGGCATTCGAGCCGCGCAACAG 796
QY 541 CAACCCCTTTGGGTCCTCCCTTGTATGGGCGCTCTGTTTCACTCTATTCTACTCT 600
Db 797 CAACCCCTTTGGGTCCTCCCTTGTATGGGAGCTCTGTTTCACTCTATTCTACTCT 856
QY 601 ATTAAATCTGCAACTGAA 619
Db 857 ATTAAATCATGCAACTGCA 875

RESULT 8
AAV29704
ID AAV29704 standard; DNA; 1329 BP.
XX
XX AAV29704;
XX 17-OCT-2003 (revised)
XX 27-AUG-2003 (revised)
XX 08-JUN-1999 (first entry)
XX
XX Clone 5M6 from MSRV-1.
XX Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
XX rheumatoid polyarthritis; ss.
XX Viruses.
XX FR2765588-A1.
XX 08-JAN-1999.
XX 07-JUL-1997; 97FR-00008816.
XX 07-JUL-1997; 97FR-00008816.
XX (INMR) BIO MERIEUX.
XX PA

```
XX WPI; 1999-098275/09.
DR P-PSDB; AAW99554.
XX
XX Nucleic acid sequences of retrovirus called MSRV-1 - associated with
PT multiple sclerosis or rheumatoid polyarthritis.
PT
XX Claim 1; Page 39-40; 83pp; French.
XX
XX This sequence represents clone 5M6 from a novel multiple sclerosis
CC related virus type 1 (MSRV1). The sequence can be used in diagnostic,
CC prophylactic or therapeutic compositions to inhibit expression of a
CC multiple sclerosis related virus and/or virus associated with rheumatoid
CC polyarthritis. (Updated on 27-AUG-2003 to correct OS field.) (Updated on
CC 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 0 U; 3 Other;
Query Match 85.8%; Score 545; DB 2; Length 1329;
Best Local Similarity 92.4%; Pred. No. 1.3e-158;
Matches 572; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 1 CCTGTATCTTTAACTCTCTTGTAAAGTTTGTCTCTTCCAGAAATCAAACTGTAAACTA 60
DB 257 CCTGTATCTTCAACTCTCTTGTAAAGTTTGTCTCTTCCAGAAATGAACTGTAAAGCTA 316
QY 61 CAAATTTGTTCTTCAATFGGAGCACCAGATGGAGTCCATGACTAAATCTACCGTGGACC 120
DB 317 CAAATAGTTCTTCAATGGAAACCCAGATGCAGTCCATGACTAAATCTACCGTGGACC 376
QY 121 CTGACCGGCTGTAGCCATGCTCCGATGTTAATGACATTTGAAGCACCCTCCCGAG 180
DB 377 CTGACCGGCTGTAGATGCTCTGATGTTAATGACATTTGAAGTCACTCCCTCCCGAG 436
QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGACGGT 240
DB 437 GAAATCTCAACTGCACAAACCCCTACTATGCCCCAATTCAGTAGGAAGCAGTTAGACAGT 496
QY 241 CATCAGCAACCTCCCAACAGCAGCTTGGTTTTCTGTGTAGAGGGGGAGCTGAGAGAC 300
DB 497 TGTACGCAACCTCCCAACAGCAGTCTGGGTTTTCTGTGTAGAGGGGGAGCTGAGAGAC 556
QY 301 AGGACTAGCTGGATTTCTTAGGCCAACAGAAATCCCTAAGCTAGCTGGGAAGGTGACT 360
DB 557 AGGACTAGCTGGATTTCTTAGGCTGACTAAGATCCCNAGCCCTNCTGGAAGGTGACC 616
QY 361 GCATCCACCTTAAACATGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
DB 617 GCATCCATCTTAAACATGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 676
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCTG 480
DB 677 ACTAAATGCTAATCAGCAAAACAGAGGTAAAGCAATAGCAATCATCTATTGCTG 736
QY 481 AGAGCAGCGGAGGAGCAAGAGTCGGATATAAACCAGGATTCAGCCCGCAACGG 540
DB 737 AGAGCAGCGGAGGAGCAAGAGTTGGATATAAACCAGGATTCAGGCAATCAAGCCAGCAACAG 796
QY 541 CAACCCCTTTGGTCCCTCCCTCTTGTATGGGGCTCTGTTTCACTCTATTTCACCTCT 600
DB 797 CAACCCCTTTGGTCCCTCCCTCCCAATGTATGGGAGCTCTGTTTCACTCTATTTCACCTCT 856
QY 601 ATTAAATCTTGCAACTGAA 619
DB 857 ATTAAATCATGCAACTGCA 875
RESULT 9
ADG14849
ID ADG14849 standard; cDNA; 1329 BP.
XX
AC ADG14849;
XX
```

```
DT 26-FEB-2004 (first entry)
XX
DE MSRV associated cDNA #11.
XX
KW ss; pol gene; retrovirus; multiple sclerosis; rheumatoid arthritis.
XX
OS Multiple sclerosis associated retrovirus.
XX
PN US2003198647-A1.
XX
PD 23-OCT-2003.
XX
PF 03-APR-2002; 2002US-00114104.
XX
PR 26-NOV-1996; 96US-00756429.
XX
PR 26-NOV-1997; 97US-00979847.
XX
PA (INMR ) BIO MERIEUX.
XX
PI Perron H, Beseme F, Bedin F, Paranhos-Baccala G;
PI Komurian-Pradel F, Jollivet-Reynaud C, Mandrand B, Garson JA, Tuke PW;
XX WPI; 2004-032461/03.
XX
XX New isolated nucleic acid and their fragments having the pol gene of a
PT retrovirus, useful for diagnosing, preventing and/or treating multiple
PT sclerosis and/or rheumatoid arthritis.
XX
PS Disclosure; SEQ ID NO 108; 193pp; English.
XX
CC The invention relates to an isolated nucleic acid which comprises the pol
CC gene of a retrovirus associated with multiple sclerosis or rheumatoid
CC arthritis. The methods and compositions of the present invention are
CC useful for diagnosing, preventing and/or treating multiple sclerosis
CC and/or rheumatoid arthritis. The present sequence is used in the
CC exemplification of the invention.
XX
SQ Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 0 U; 3 Other;
Query Match 85.8%; Score 545; DB 12; Length 1329;
Best Local Similarity 92.4%; Pred. No. 1.3e-158;
Matches 572; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 1 CCTGTATCTTTAACTCTCTTGTAAAGTTTGTCTCTTCCAGAAATCAAACTGTAAACTA 60
DB 257 CCTGTATCTTCAACTCTCTTGTAAAGTTTGTCTCTTCCAGAAATGAACTGTAAAGCTA 316
QY 61 CAAATTTGTTCTTCAATFGGAGCACCAGATGGAGTCCATGACTAAATCTACCGTGGACC 120
DB 317 CAAATAGTTCTTCAATGGAAACCCAGATGCAGTCCATGACTAAATCTACCGTGGACC 376
QY 121 CTGACCGGCTGTAGCCATGCTCCGATGTTAATGACATTTGAAGCACCCTCCCGAG 180
DB 377 CTGACCGGCTGTAGATGCTCTGATGTTAATGACATTTGAAGTCACTCCCTCCCGAG 436
QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGACGGT 240
DB 437 GAAATCTCAACTGCACAAACCCCTACTATGCCCCAATTCAGTAGGAAGCAGTTAGACAGT 496
QY 241 CATCAGCAACCTCCCAACAGCAGCTTGGTTTTCTGTGTAGAGGGGGAGCTGAGAGAC 300
DB 497 TGTACGCAACCTCCCAACAGCAGTCTGGGTTTTCTGTGTAGAGGGGGAGCTGAGAGAC 556
QY 301 AGGACTAGCTGGATTTCTTAGGCCAACAGAAATCCCTAAGCTAGCTGGGAAGGTGACT 360
DB 557 AGGACTAGCTGGATTTCTTAGGCTGACTAAGATCCCNAGCCCTNCTGGAAGGTGACC 616
QY 361 GCATCCACCTTAAACATGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
DB 617 GCATCCATCTTAAACATGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 676
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCTG 480
```

Db 677 ACTAAATGCTAATCAGGCAAAACAGGAGGTAAAGCAATAGCAATCATCTATTGCGTG 736
QY 481 AGAGCACAGCGGAGGACAAAGATCGGATATATAAACCAGGCAATTCGAGCCGCAACGG 540
Db 737 AGAGCACAGCGGAGGACAAAGATTCGGATATAAATCAAGCAATTCAGGCAACAG 796
QY 541 CAACCCCTTTGGGTCCTCCCTTTGATGGGCGCTCTGTTTCACTCTATTTCACCTCT 600
Db 797 CAACCCCTTTGGGTCCTCCCTTTGATGGGAGCTCTGTTTCACTCTATTTCACCTCT 856
QY 601 ATTAAATCTTGCAACTGAA 619
Db 857 ATTAAATCATGCAACTGCA 875

RESULT 10
ACC46747
ID ACC46747 standard; cDNA; 2046 BP.
XX AC ACC46747;
XX 02-JUN-2003 (first entry)
XX Human dithp growth/development-associated protein-encoding cDNA.
XX Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
KW cancer; cell proliferative disorder; autoimmune disorder;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging; growth;
KW development; gene; ss.
XX Homo sapiens.
XX W0200297031-A2.
PN 05-DEC-2002.
XX 27-MAR-2002; 2002WO-US010056.
XX 28-MAR-2001; 2001US-0279619P.
PR 29-MAR-2001; 2001US-0280067P.
PR 29-MAR-2001; 2001US-0280068P.
PR 16-MAY-2001; 2001US-0291280P.
PR 17-MAY-2001; 2001US-0291829P.
PR 19-MAY-2001; 2001US-0291849P.
PR 19-JUN-2001; 2001US-0299428P.
PR 20-JUN-2001; 2001US-0299776P.
PR 20-JUN-2001; 2001US-0300001P.
XX (INCY-) INCYTE GENOMICS INC.
PA Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
XX Dufour GB, Hillman JL, Yu JY, Tuason O, Yap PE, Anshay SR;
PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI Flores V, Marwaha R, Lo A, Lan RV, Urashka ME;
XX WPI; 2003-129518/12.
DR P-PSDB; ABR41810.
XX Novel human diagnostic and therapeutic polypeptide useful for identifying
PT test compound which specifically binds to a polypeptide encoded by human
PT diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX Claim 2; SEQ ID NO 668; 591pp; English.
XX The invention relates to novel human diagnostic and therapeutic
XX polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
CC proteins (DITHP; ABR41136-ABR41812). The invention also relates to
CC polynucleotide sequences at least 90% identical to the dithp cDNA

CC sequences of the invention; recombinant vectors, host cells and
CC transgenic organisms comprising a dithp nucleic acid sequence; the
CC recombinant production of DITHP proteins; antibodies specific for DITHP
CC proteins; microarrays comprising dithp nucleic acid sequences; methods of
CC detecting dithp nucleotide and protein sequences; methods of screening
CC for compounds which specifically bind a DITHP protein; and methods of
CC assessing the toxicity of test compounds using a dithp hybridisation
CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
CC diagnosis of a wide variety of conditions including cancer and other cell
CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,
CC viral, fungal or parasitic infections; hormonal disorders; metabolic
CC disorders; neurological disorders; gastrointestinal disorders; transport
CC disorders; and connective tissue disorders. They may also be used to
CC screen for modulators of protein activity or gene expression. DITHP
CC proteins can additionally be used in analysis of the proteome of a tissue
CC or cell type and to induce antibodies. The dithp nucleic acids are
CC additionally useful in somatic or germline gene therapy of the disorders
CC mentioned above, as a source of antisense sequences, as a source of
CC probes and primers, in genotyping and identification of individuals, in
CC the generation of transgenic animal models of human disease or knock in
CC humanised animals, in toxicological testing, and in transcript imaging.
CC The present sequence represents a dithp cDNA encoding a DITHP protein
CC which is associated with growth and development. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2046 BP; 587 A; 494 C; 391 G; 566 T; 0 U; 8 Other;

Query Match 85.1%; Score 540.2; DB 8; Length 2046;
Best Local Similarity 94.4%; Pred. No. 5e-157;
Matches 560; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1 CCTGTATCTTTAACTCTCTTGTAAAGTTTGTCTCTCCAGATCAAACTGTAACATA 60
Db 571 CTCTGTATCTTTAACTCTCTTGTAAAGTTTGTCTCTCCAGATCAAACTGTAACATA 630
QY 61 CAAATGTTCTTCAAAATGGAGCACAGATGGAGTCCATGACTAAGATCCACCGTGACCC 120
Db 631 CAAATGTTCTTCAAAATGGAGCCCCAGATGCGATGCGATCAAGATCCACCAAGACCC 690
QY 121 CTGGACCGCTCTAGCCCATGCTCCGATGTTAATGACATTAAGAGCAACCTCCCGAG 180
Db 691 CTGGACCGCTCTAGCCCATGCTCCGATGTTAATGACATTAAGAGCAACCTCCCGAG 750
QY 181 GAAATCTCAACTGCAACACCTCTACTATGCCCCAATTCAGCGGAGAGCTAGACGGT 240
Db 751 GAAATCTCAACTGTCACACCTCTACTATGCCCCAATTCAGCGGAGAGCTAGACGGT 810
QY 241 CATCAGCCAACTCTCCCAACAGACATTTGGGTTTCTGTTGAGAGGGGAGCTGAGAGAC 300
Db 811 CATCAGCCAACTCTCCCAACAGACATTTGGGTTTCTGTTGAGAGGGGAGCTGAGAGAC 870
QY 301 AGACTAGCTGGATTTCTTAGCCCAACAGAGATCCCTAAGCCTAGCTGGGAAGGTGACT 360
Db 871 AGACTAGCTGGATTTCTTAGCCCGACTAAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 930
QY 361 GCATCCACCTCTAAACATGGGCTTGCAACTTAGCTCACAACCCGACCAATCAGAGAGCTC 420
Db 931 GCATCCACCTTTAAACATGGGCTTGCCACTTAGCTCACAACCCGACCAATCAGAGAGCTC 990
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTTAAGAAATAGCAATCATCTATTGCGTG 480
Db 991 ACTAAATGCTAATTAGGCAAAACAGGAGGTAAAGAAATAGCAATCATCTATTGCGTG 1050
QY 481 AGAGCACAGCGGAGGAGCAAGGATCGGATATAAACCAGGCAATTCGAGCCGCAACGG 540
Db 1051 AGAGCACAGCGGAGGAGCAAGGATCGGATATAAACCAGGCAATTCGAGCCGCAATGG 1110
QY 541 CAACCCCTTTGGGTCCTCCCTTTGATGGGCGCTCTGTTTCACTCTATT 593
Db 1111 CAACCCCTTTGGGTCCTCCCTTTGATGGGAGCTCTGTTTCACTCTATT 1163

RESULT 11

ADB84403
ID ADB84403 standard; DNA; 1329 BP.

XX ADB84403;
XX AC
XX DT 04-DEC-2003 (first entry)
XX MSRV-1 associated DNA sequence #21.
DE ds; multiple sclerosis; rheumatoid arthritis; gag; pol;
XX reverse transcriptase; ribonuclease H.
KW Unidentified;
XX OS

XX US2003039664-A1.

XX 27-FEB-2003.

XX 26-NOV-1997; 97US-00979847.

XX 26-NOV-1996; 96US-00756429.

XX (PERR/) PERRON H.

XX (BESE/) BESEME F.

XX (BEDI/) BEDIN F.

XX (PARA/) PARANHOS-BACCALA G.

XX (KOMU/) KOMURIAN-PRADEL F.

XX (JOLI/) JOLIVET-REYNAUD C.

XX (MAND/) MANDRAND B.

XX (GARS/) GARSON J A.

XX (TUKE/) TUKE P W.

XX Perron H, Beseme F, Bedin F, Paranhos-Baccala G;

PI Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B, Garson JA, Tuke PW;

XX WPI; 2003-512253/48.

XX New isolated or purified nucleic acid associated with multiple sclerosis
PT and/or rheumatoid arthritis, useful for detecting a virus associated with
PT multiple sclerosis or rheumatoid arthritis in a biological sample.

XX Claim 31; Page 80; 193pp; English.

XX The invention relates to an isolated or purified nucleic acid from a
CC virus associated with multiple sclerosis and/or rheumatoid arthritis.
CC multiple sclerosis-associated virus (MSRV)-1. The nucleic acids comprise
CC pol, gag or reverse transcriptase genes (or their fragments) encoding the
CC proteins or defined peptides (including immunodominant peptides,
CC antigenic peptides or conserved motifs). Also included are a process for
CC detecting a virus associated with multiple sclerosis or rheumatoid
CC arthritis in a biological sample, a nucleic acid probe for the detection
CC of a virus associated with multiple sclerosis or rheumatoid arthritis, a
CC primer for the amplification by polymerisation of a nucleic acid of a
CC viral material associated with multiple sclerosis or rheumatoid
CC arthritis, a polypeptide exhibiting an inhibitory activity on the
CC proteolytic, reverse transcriptase or ribonuclease H activity from MSRV,
CC and an antibody directed against the MSRV-1 virus obtained by
CC immunologically reacting a human or animal body or cells with an
CC immunogenic agent consisting of the antigenic polypeptide defined above.
CC The nucleic acids are useful for detecting a biological sample a virus
CC associated with multiple sclerosis or rheumatoid arthritis, or for
CC detecting in a biological sample, the presence of or exposure to a virus
CC associated with multiple sclerosis or rheumatoid arthritis. The present
CC sequence is a claimed MSRV-associated sequence whose identity cannot be
CC accurately determined. Note: The SEQ ID numbers for the sequences as
CC displayed in the main body of the patent do not match the SEQ ID numbers
CC in the sequence listing. Consequently those sequences mentioned in the
CC claims may not be the sequences the authors intended to claim.

XX Sequence 1329 BP; 339 A; 346 C; 282 G; 332 T; 0 U; 30 Other;

Query Match 84.3%; Score 535; DB 9; Length 1329;

Best Local Similarity 90.8%; Pred. No. 1.7e-155;
Matches 562; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY	1	CCCTGTATCTTTAACTCCCTTGTAAAGTTTGTCTCTTCCAGAATCAAACTGTAAAACTA	60
DB	257	CCCTGTATCTTCAACTTCTTGTAAAGTTTGTCTCTTCCAGAANTGAGCTGTAAAGCTA	316
QY	61	CAAAATTGTTCTTCAAAATGGAGCACCAAGATGAGTCCATGATCAAGATCCACCTGGACCC	120
DB	317	CAAAATAGTTCTTCAAAATGGAAACCCAGATGACGTCCATGACTANAATCTACCGTGGACCC	376
QY	121	CTGGACCGGCTGCTAGCCCATGCTCCGATGTTAATGACATTCGAGGACACCTCCCGAG	180
DB	377	CTGGACCGGCTGCTAGACTATGCTCTGATGTTAATGACATTCGAGGACACCTCCCGAG	436
QY	181	GAATCTCAACTGCACAAACCCCTACTATGCCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT	240
DB	437	GAATCTCAACTGCACAAACCCCTACTACTCCAAATTCAGTAGNAGCAGTTAGAGCAGT	496
QY	241	CATCAGCCAACTCCCAACAGCACCTTGGGTTTTCTGTGAGAGGGGGGAGTGGAGAGAC	300
DB	497	TGTCAGCCAACTCCCAACAGCACCTTGGGTTTTCTGTGAGAGGGGGTGGAGAGAC	556
QY	301	AGGACTAGCTGGATTTCTTAGCCCAACGAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACT	360
DB	557	AGGACTAGCTGGATTTCTTAGCCCAACGAGAAATCCCAAGCCCAATTCAGGAGGTGACC	616
QY	361	GCATCCACCTCTAAACATGCGGCTTCCAACTTAGCTCACACCCGACCAATCAGAGAGCTC	420
DB	617	GCATCCATCTTTAAACATGCGGCTTCCAACTTAGCTCACACCCGACCAATCAGAGAGCTC	676
QY	421	ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCTG	480
DB	677	ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGCAATAGCAATCATCTATTGCTG	736
QY	481	AGAGCAGCGGGGAGGACGAGATCGGATATAAACCAGGATTCGAGCGGCAACGG	540
DB	737	AGAGCAGCGGGGAGGACGAGATTCGATATAAACCAGGATTCGAGCGGCAACAG	796
QY	541	CAACCCCTTTGGGTCCCTCCCTTTGATGGCGCTCTGTTTTCATCTATTTCATCTCT	600
DB	797	CAACCCCTTTGGGTCCCTCCCTTTGATGGAGCTCTGTTTTCATCTATTTCATCTCT	856
QY	601	ATTAAATCTTGCAACTGAA	619
DB	857	ATTAAATCATGCAACTGCA	875

RESULT 12
AAX77526
ID AAX77526 standard; cDNA; 2946 BP.
XX
XX AAX77526;
AC
XX
DT 10-AUG-1999 (first entry)
XX
DE Human secreted protein A172_2 cDNA.
XX
KW Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
KW bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
KW cell proliferation; cell differentiation; suppressor; tumour inhibition;
KW haematopoiesis regulator; activin; inhibitor; chemotactic; chemokinetic;
KW haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
KW cadherin; tumour invasion suppressor; gene therapy; tissue growth; BS.
XX
OS Homo sapiens.
XX
PN WO9926972-A1.
XX
PD 03-JUN-1999.
XX
PF 17-NOV-1998; 98WO-US024614.

XX 21-NOV-1997; 97US-00976110.
PR 18-MAY-1998; 98US-00080478.
PR 20-OCT-1998; 98US-00175928.
XX (GEMY) GENETICS INST INC.
PA
XX Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M;
XX WPI; 1999-357813/30.
DR P-PSDB; AAY08622.
XX
PT New polynucleotides encoding secreted proteins.
PS Claim 13a; Page 100-101; 142pp; English.
XX
CC This invention describes novel human secreted proteins encoded by
CC polynucleotides isolated from human adult testes, adult brain, adult
CC blood or adult placenta, or murine adult bone marrow or thymus cDNA
CC libraries. The products of the invention are predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals, although no
CC supporting data is given. Suggested activities include nutritional
CC activity, cytokine and cell proliferation/differentiation activity,
CC immune stimulating (e.g. as vaccines) or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC actin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, and tumour
CC inhibition activity. The polynucleotides are also stated to be useful for
CC gene therapy
XX
SQ Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 U; 0 Other;

Query Match 84.1%; Score 533.8; DB 2; Length 2946;
Best Local Similarity 91.5%; Pred. No. 5.9e-155;
Matches 581; Conservative 0; Mismatches 42; Indels 12; Gaps 1;

QY 1 CCTGTATCTTTAACTCCCTTGTAGTTTGTCTTCCAGAACTCAAACTGTAAACTA 60
DB 2323 CCTGTATCTTTAACTCCCTTGTAGTTTGTCTTCCAGAACTCAAACTGTAAACTA 2382

QY 61 CAAATGTTCTTCAATGGAGCACCAGATGAGTCCATGACTAAGATCCACCGTGACCC 120
DB 2383 -----CAAATGGAGCACCAGATGAGTCCAGAACTAAGATCTACCGCAGACCC 2430

QY 121 CTGACCGGCTGCTAGCCATGCTCCGATGTTAATGACATTCAGGCACCCCTCCGAG 180
DB 2431 CTGACCGGCTGCTAGCCATGCTGTTAATGACATTCAGGCACCCCTCCGAG 2490

QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAAGCAGTTAGACGGT 240
DB 2491 GAAATCTCACTGCACAACTCTACTAGCCGCCCAATTCAGCAGGAGCAGTTAGACGGT 2550

QY 241 CATCAGCAACCTCCCAACAGCACTTGGTTTTCTGTGTTGAGAGGGGGAGTCAAGAC 300
DB 2551 CGTCGGCAACCTCCCAACAGCACTTGGTTTTCTGTGTTGAGAGGGGGAGTCAAGAC 2610

QY 301 AGACTAGTGGATTTCTTAGGCCAACGAGATCCCTTAGCCCTAGCTAGTGGGAAGTGACT 360
DB 2611 AGACTAGTGGATTTCTTAGGGTGAATAAGATCCCTTAGCCCTAGCTAGTGGGAAGTGACC 2670

QY 361 GCATCCACCTTAAACATGGGGTTGCACTTAGCTCACAACCCGACCAATCAGAGAGCTC 420
DB 2671 ACATCCACCTTAAACAGGGGTTGCACTTAGCTCACAACCTGACCAATCAGAGAGCTC 2730

QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAGAAATAGCAATCATCTATTGCGTG 480
DB 2731 ACTAAATGCTAATTAGGCAAAACAGGAGTAAGAAATAGCAATCATCTATTGCGTG 2790

QY 481 AGACACAGCGGAGGACAAAGATCCGGATATAAACCCAGGCATTCGAGCGGCAACGG 540
|||||

DB 2791 AGAGCACAGCAGGAGGACAATGATCGGATATAACCAACTCTTCGAGCGGCAACGG 2850
QY 541 CAACCCCTTTGGTCCCTCCCTTGTATGGGCGCTGTGTTTCACTTATTCACTCT 600
|||||
DB 2851 CAACCCCTTTGGTCCCTCCCTTGTATGGGAGCTGTGTTTCACTTATTCACTCT 2910
|||||
QY 601 ATTAATCTTGGCACTGAAAAAAGAAAAA 635
|||||
DB 2911 ATTAATCTTGGCACTGAAAAAAGAAAAA 2945
|||||

RESULT 13
AAZ59468
ID AAZ59468 standard; cDNA; 2946 BP.
XX
XX AAZ59468;
AC
XX 11-APR-2000 (first entry)
DT Human secreted protein AJ172_2 polynucleotide sequence.
XX
DE Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;
XX placental pathology; metastasis inhibition; nutritional activity;
KW immune stimulator; haematopoiesis regulator; tissue growth;
KW tumour inhibitor; anti-inflammatory; clone AJ172_2; ATCC_98115;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
XX WO9960020-A1.
PN 25-NOV-1999.
XX
PD 17-MAY-1999; 99WO-US010915.
PF
XX 18-MAY-1998; 98US-00080478.
PR 20-OCT-1998; 98US-00175928.
XX
XX (GEMY) GENETICS INST INC.
XX
XX Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Mi S, Treacy M;
PI
XX WPI; 2000-116311/10.
DR P-PSDB; AAY67313.
XX
XX New polynucleotides encoding secreted cDNA libraries, used to develop
PT products for the diagnosis and treatment of neoplastic disease.
XX
PS Claim 14; Page 107-108; 149pp; English.
XX
CC This is the human secreted protein AJ172_2 nucleotide sequence, obtained
CC from a human adult testes cDNA library. The invention relates to secreted
CC human and murine proteins. The polynucleotides and proteins are predicted
CC to have biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals. Detection of the levels of the proteins can be used for the
CC diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents
CC which modulate the expression or function of the proteins may be used for
CC treating a neoplastic disease and inhibiting metastasis. Other suggested
CC activities include nutritional activity (e.g. in feeds), cytokine and
CC cell proliferation/differentiation activity, haematopoiesis regulating activity,
CC tissue growth activity, actin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity. The
CC polynucleotide sequences are also stated to be useful for gene therapy
XX
SQ Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 U; 0 Other;

Query Match 84.1%; Score 533.8; DB 3; Length 2946;
Best Local Similarity 91.5%; Pred. No. 5.9e-155;

Matches 581; Conservative 0; Mismatches 42; Indels 12; Gaps 1;

Qy 1 CCCTGTATCTTAACTCCCTTGTAACTTTGTCTCTCCAGAAATCAAACTGTAATACTA 60
Db 2323 CCCTGTATCTTAACTCCCTTGTAACTTTGTCTCTCCAGAAATCAAACTGTAATACTA 2382

Qy 61 CAAATTTGTTCTTCAAAATGGAGACACAGATGAGTCCATGACTAAGATCCACCGTGGACCC 120
Db 2383 -----CAAATGGAGCCCAAGATGAGTCCAAAGACTAAGATCTTACCGCAGACCC 2430

Qy 121 CTGACCGGCTGCTAGTCCCATGCTCCGATGTTAATGACATTTGAAGGCACCCCTCCCGAG 180
Db 2431 CTGACCGGCTGCTAGTCCCATGCTCCGATGTTAATGACATTTGAAGGCACCCCTCCCGAG 2490

Qy 181 GAAATCTCAATGACCAACCCCTACTATGCCCCAATTCAGCGGGAACAGTTAGAGCGT 240
Db 2491 GAAATCTCAGTCAGACCAACCTCTACTACGCCCCAATTCAGCGGGAACAGTTAGAGCGT 2550

Qy 241 CATCAGCAACCTCCCAACAGCACTTGGGTTTTCTGTTGAGAGGGGACCTGAGAGAC 300
Db 2551 CGTCGGCAACCTCCCAACAGCACTTGGGTTTTCTGTTGAGATGGGGACCTGAGAGAC 2610

Qy 301 AGGACTAGCTGGATTTCTTAGGCCAACGAAAGATCCCTAAGCCCTAGCTGGGAAAGTGACT 360
Db 2611 AGGACTAGCTGGATTTCTTAGGCTGACTAAGATCCCTAAGCCCTAGCTGGGAAAGTGACC 2670

Qy 361 GCATCCACCTTAAACATGGGGCTTGCACTTGTAGTCAACCCGACCAATCAGAGACTC 420
Db 2671 ACATCCACCTTAAACATGGGGCTTGCACTTGTAGTCAACCCGACCAATCAGAGACTC 2730

Qy 421 ACTAAATGCTAATAGGCAAAATAGGAGTAAGAAATAGCAATCATCTATTCGCTG 480
Db 2731 ACTAAATGCTAATAGGCAAAATAGGAGTAAGAAATAGCAATCATCTATTCGCTG 2790

Qy 481 AGAGCAGCGGGAGGACAAAGATCGGGATATAAACCAGGCAATTCAGCCGCGCAACGG 540
Db 2791 AGAGCAGCGGGAGGACAAATGATCGGGATATAAACCAGGCAATTCAGCCGCGCAACGG 2850

Qy 541 CAACCCCTTTGGTCCCTCCCTTTGATGGGCTCTGTTTCACTATTTCACTCT 600
Db 2851 CAACCCCTTTGGTCCCTCCCTTTGATGGGCTCTGTTTCACTATTTCACTCT 2910

Qy 601 ATTAAATCTTGCACTGAAAAA 635
Db 2911 ATTAAATCTTGCACTGAAAAA 2945

RESULT 14
ADC38776
ID ADC38776 standard; cDNA; 2946 BP.
XX
AC ADC38776;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human cDNA encoding a secreted protein #63.
XX
KW ss; gene; immune disorder; severe combined immunodeficiency; SCID;
KW autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
KW rheumatoid arthritis; allergic reaction; asthma; myeloid cell deficiency;
KW lymphoid cell deficiency; osteoporosis; osteoarthritis;
KW peripheral nervous system disease; peripheral neuropathy;
KW Alzheimer's disease; Parkinson's disease; coagulation disorder;
KW inflammatory disease; systemic inflammatory response syndrome; SIRS;
KW ischaemia-reperfusion injury; Crohn's disease; anaphylaxis;
KW hypersensitivity; regeneration; neural cell proliferation; fertility;
KW tumour; chemokine; human; secreted protein.
XX
OS Homo sapiens.
XX
PN US2002193567-A1.
XX
PD 19-DEC-2002.

XX 02-APR-2002; 2002US-00114893.
XX 11-AUG-1995; 95US-00514014.
PR 05-APR-1996; 96US-00628364.
PR 19-APR-1996; 96US-00635311.
PR 07-JUN-1996; 96US-00659224.
PR 17-JUN-1996; 96US-00664596.
PR 09-JUL-1996; 96US-00677231.
PR 26-JUL-1996; 96US-00686878.
PR 23-AUG-1996; 96US-00701819.
PR 27-SEP-1996; 96US-00721488.
PR 27-SEP-1996; 96US-00721798.
PR 27-SEP-1996; 96US-00721923.
PR 27-SEP-1996; 96US-00721926.
PR 25-OCT-1996; 96US-00738367.
PR 30-OCT-1996; 96US-00739775.
PR 13-JAN-1997; 97US-00783395.
PR 10-APR-1997; 97US-00833823.
PR 02-JUN-1997; 97US-00867677.
PR 05-SEP-1997; 97US-00924838.
PR 06-OCT-1999; 99US-00413232.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Bowman MR, Spaulding V, Carlin-Duckett M;
PI Kelleher K;
XX
XX WPI: 2003-657236/62.
DR P-PSDB; ADC38777.
XX
PT Proteins AZ3021 encoded by clone AZ3021 from human adult colon, and
PT BD12716 encoded by clone BD12716 from human fetal kidney cDNA library,
XX useful for treating e.g. multiple sclerosis and rheumatoid arthritis.
XX
XX Disclosure; SEQ ID NO 134; 412pp; English.
XX
CC The invention relates to a protein comprising fully defined A2302 1
CC protein or BD127 1 6 protein. The polynucleotides are useful for
CC expressing recombinant proteins for analysis and are also useful as
CC chromosome markers or tags to identify chromosomes or to map related gene
CC positions. The proteins are useful as amino acid supplement, carbon
CC source, nitrogen source and carbohydrate source. The proteins are useful
CC for treating various immune deficiencies and disorders (e.g. severe
CC combined immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
CC sclerosis, systemic lupus erythematosus, rheumatoid arthritis), allergic
CC reactions (e.g. asthma), myeloid or lymphoid cell deficiencies, (e.g.
CC osteoporosis or osteoarthritis, peripheral nervous system diseases (e.g.
CC peripheral neuropathy, Alzheimer's disease, Parkinson's disease),
CC coagulation disorders, inflammatory diseases (e.g. systemic inflammatory
CC response syndrome (SIRS), ischaemia-reperfusion injury, Crohn's disease),
CC anaphylaxis and hypersensitivity. Proteins are also useful for inducing
CC tumour immunity, for inducing bone, cartilage, tendon, ligament and/or
CC nerve growth or regeneration, for proliferating neural cells and for
CC regenerating nerve and brain tissue, for inducing fertility and for
CC inhibiting tumour growth. Proteins are also useful as chemokine for
CC mammalian cells (e.g., monocytes, fibroblasts, neutrophils), and also
CC useful as inhibitors of receptor/ligand interactions. The present
CC sequence represents a cDNA encoding a human secreted protein.
XX
SQ Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 U; 0 Other;
Query Match 84.1%; Score 533.8; DB 10; Length 2946;
Best Local Similarity 91.5%; Pred. No. 5.9e-155;
Matches 581; Conservative 0; Mismatches 42; Indels 12; Gaps 1;
Qy 1 CCCTGTATCTTAACTCCCTTGTAACTTTGTCTCTCCAGAAATCAAACTGTAATACTA 60
Db 2323 CCCTGTATCTTAACTCCCTTGTAACTTTGTCTCTCCAGAAATCAAACTGTAATACTA 2382
Qy 61 CAAATTTGTTCTTCAAAATGGAGACCAAGATGAGTCCATGACTAAGATCCACCGTGGACCC 120
Db 2383 CCCTGTATCTTAACTCCCTTGTAACTTTGTCTCTCCAGAAATCAAACTGTAATACTA 2382
Qy 61 CAAATTTGTTCTTCAAAATGGAGACCAAGATGAGTCCATGACTAAGATCCACCGTGGACCC 120
Db 2383 CCCTGTATCTTCAAAATGGAGACCAAGATGAGTCCATGACTAAGATCCACCGTGGACCC 120

Db	2383	-----CAAATGGAGCCCAAGATGCACTCCAAAGACTAAGATCTACCCAGACCC	2430	FT	intron	/number= 1
Qy	121	CTGACCGGCTGCTAGCCCATGCTCGATGTTAATGACATTAAGAGCACCCTCCCGAG	180	FT	FT	3170..53737
Db	2431	CTGACCGGCTGCTAGCCCAAGATCTGATGTTAATGACATTAAGAGCACCCTCCCGAG	2490	FT	variation	/*tag= c
Qy	181	GAATCTCAACTGCACAAACCCCTACTATGCCCCCAATTCAGCGGAAGCAGTTAGACGGT	240	FT	variation	replace(3180,G)
Db	2491	GAATCTCACTGCACAACTCTACTAGCGCCCCCAATTCAGCGGAAGCAGTTAGACGGT	2550	FT	FT	/tag= am
Qy	241	CATCAGCAACCTCCCAACAGACACTGGTGTTCCTGTTGAGAGGGGACTGAGAGAC	300	FT	variation	/standard_name= "Single nucleotide polymorphism"
Db	2551	CGTCGGCAACCTCCCAACAGACACTTAGGTTTTCCTGTTGAGATGGGGACTGAGAGAC	2610	FT	FT	replace(7590,A)
Qy	301	AGACTAGCTGGATTTCTTAGGCCAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT	360	FT	variation	/*tag= an
Db	2611	AGACTAGCTGGATTTCTTAGGCTGACTAAGAATCCCTAAGCCTAGCTGGGAAGGTGACC	2670	FT	variation	replace(9315,C)
Qy	361	GCATCCACCTCTAAACATAGGGGCTTGCACTTAGCTCACACCCGACCAATCAGAGAGCTC	420	FT	variation	/tag= ao
Db	2671	ACATCCACCTTTAAACACGGGCTTGCACTTAGCTCACACCTGACCAATCAGAGAGCTC	2730	FT	FT	replace(9572,A)
Qy	421	ACTAAATGCTAATAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCTTG	480	FT	variation	/*tag= ap
Db	2731	ACTAAATGCTAATAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCTTG	2790	FT	variation	/standard_name= "Single nucleotide polymorphism"
Qy	481	AGAGCAGCGGGAGGACAGGATCGGATATAAACCCAGGCAATTCGAGCCGGCAACGG	540	FT	FT	replace(11004,G)
Db	2791	AGAGCAGCGGGAGGACAAATGATCGGATATAAACCCAGGCAATTCGAGCCGGCAACGG	2850	FT	variation	/tag= as
Qy	541	CAACCCCTTTGGTCCCTCCCTTTGATGGCGCTCTGTTTCACTCTATTTCATCTCT	600	FT	variation	replace(12478,A)
Db	2851	CAACCCCTTTGGTCCCTCCCTTTGATGGAGCTCTGTTTCACTCTATTTCATCTCT	2910	FT	FT	/standard_name= "Single nucleotide polymorphism"
Qy	601	ATTAATCTTGCACTGAAAAAATAAAAAA 635		FT	variation	/*tag= at
Db	2911	ATTAATCTTGCACTGAAAAAATAAAAAA 2945		FT	FT	/standard_name= "Single nucleotide polymorphism"
RESULT 15					variation	replace(13904,G)
ABK89296/c					FT	/tag= au
ID ABK89296 standard; DNA; 326014 BP.					FT	/standard_name= "Single nucleotide polymorphism"
XX	AC	ABK89296;		FT	variation	replace(25444..25446,TT)
XX	DT	16-OCT-2002 (first entry)		FT	FT	/*tag= av
XX	DE	Human gene for novel serine/threonine serine kinase.		FT	variation	/standard_name= "Single nucleotide polymorphism"
XX	KW	Human; ds; gene; serine/threonine protein kinase; inflammation; cancer;		FT	FT	replace(26805,A)
KW	KW	arteriosclerosis; psoriasis; SNF kinase; transgenic; chromosome 3; SNP;		FT	variation	/tag= aw
KW	KW	single nucleotide polymorphism.		FT	FT	/standard_name= "Single nucleotide polymorphism"
XX	OS	Homo sapiens.		FT	variation	replace(28010,R)
XX	XX	Key	Location/Qualifiers	FT	FT	/*tag= ax
XX	XX	variation	replace(1139..1140,GGT)	FT	variation	/standard_name= "Single nucleotide polymorphism"
FT	FT	FT	/*tag= ai	FT	FT	replace(28107,Y)
FT	FT	FT	/standard_name= "Single nucleotide polymorphism"	FT	variation	/*tag= ay
FT	FT	FT	replace(1399,G)	FT	FT	/standard_name= "Single nucleotide polymorphism"
FT	FT	FT	/*tag= aj	FT	variation	replace(28178,R)
FT	FT	FT	/standard_name= "Single nucleotide polymorphism"	FT	FT	/*tag= az
FT	FT	FT	replace(1871..1873,TG)	FT	variation	/standard_name= "Single nucleotide polymorphism"
FT	FT	FT	/*tag= ak	FT	FT	replace(28356,T)
FT	FT	FT	/standard_name= "Single nucleotide polymorphism"	FT	FT	/tag= ba
FT	FT	FT	replace(1875..1877,TA)	FT	variation	/standard_name= "Single nucleotide polymorphism"
FT	FT	FT	/*tag= al	FT	FT	replace(29535,G)
FT	FT	FT	3000..323016	FT	variation	/*tag= bb
FT	FT	FT	/product= "Serine/threonine protein kinase"	FT	FT	/standard_name= "Single nucleotide polymorphism"
FT	FT	FT	/*tag= a	FT	variation	replace(31808,T)
FT	FT	FT	3000..3169	FT	FT	/*tag= bc
FT	FT	FT	/*tag= b	FT	FT	/standard_name= "Single nucleotide polymorphism"

FT	variation	replace(37470,G)	
FT		/tag= bj	"Single nucleotide polymorphism"
FT	variation	/standard name=	replace(37481,A)
FT		/tag= bk	"Single nucleotide polymorphism"
FT	variation	/standard name=	replace(37499,T)
FT		/tag= bl	"Single nucleotide polymorphism"
FT	variation	/standard name=	replace(40178, .40179, GRA)
FT		/tag= bm	"Single nucleotide polymorphism"
FT	variation	/standard name=	replace(40787,A)
FT		/tag= bn	"Single nucleotide polymorphism"
FT	variation	/standard name=	replace(44582,G)
FT		/tag= bo	"Single nucleotide polymorphism"
FT	variation	/standard name=	replace(45615,A)
FT		/tag= bp	"Single nucleotide polymorphism"
FT	variation	/standard name=	replace(46091,T)
FT		/tag= bq	"Single nucleotide polymorphism"
FT	variation	/standard name=	replace(47844,C)
FT		/tag= br	"Single nucleotide polymorphism"
FT	variation	/standard name=	replace(48013,T)
FT		/tag= bs	"Single nucleotide polymorphism"
FT	variation	/standard name=	replace(48072,A)
FT		/tag= bt	"Single nucleotide polymorphism"
FT	variation	/standard name=	replace(50802,T)
FT		/tag= bu	"Single nucleotide polymorphism"
FT	variation	/standard name=	replace(53305,A)
FT		/tag= bv	"Single nucleotide polymorphism"
FT	variation	/standard name=	replace(53378, .53380, GG)
FT		/tag= bw	"Single nucleotide polymorphism"
FT	exon	/standard name=	53738, .53903
FT	exon	/tag= d	
FT	intron	/number= 2	
FT		53904, .83114	
FT		/tag= e	
FT	variation	/number= 2	
FT		replace(56233,G)	
FT	variation	/tag= bx	"Single nucleotide polymorphism"
FT		/standard name=	replace(56453,C)
FT	variation	/tag= by	"Single nucleotide polymorphism"
FT		/standard name=	replace(59300,T)
FT	variation	/tag= bz	"Single nucleotide polymorphism"
FT		/standard name=	replace(60998,A)
FT	variation	/tag= ca	"Single nucleotide polymorphism"
FT		/standard name=	replace(63013,A)
FT	variation	/tag= cb	"Single nucleotide polymorphism"
FT		/standard name=	replace(63293,A)
FT	variation	/tag= cc	"Single nucleotide polymorphism"
FT		/standard name=	replace(64181,T)
FT	variation	/tag= cd	"Single nucleotide polymorphism"
FT		/standard name=	replace(64191,A)
FT	variation	/tag= ce	"Single nucleotide polymorphism"
FT		/standard name=	replace(68992,C)
FT	variation		

FT	/*tag= cf	/standard_name= "Single nucleotide polymorphism"	
FT	replace(76246,Y)	/*tag= cg	
FT	variation	/standard_name= "Single nucleotide polymorphism"	
FT	replace(77368,G)	/*tag= ch	
FT	variation	/standard_name= "Single nucleotide polymorphism"	
FT	replace(78043,A)	/*tag= ci	
FT	variation	/standard_name= "Single nucleotide polymorphism"	
FT	replace(79395,T)	/*tag= cj	
FT	variation	/standard_name= "Single nucleotide polymorphism"	
FT	replace(79474,G)	/*tag= ck	
FT	variation	/standard_name= "Single nucleotide polymorphism"	
FT	replace(79992,T)	/*tag= cl	
FT	variation	/standard_name= "Single nucleotide polymorphism"	
FT	replace(81065,C)	/*tag= cm	
FT	variation		
Query Match	83.9%;	Score 532.8;	DB 6; Length 326014;
Best Local Similarity	90.5%;	Pred. No. 1.2e-153;	
Matches 583;	Conservative 0;	Mismatches 52;	Indels 9; Gaps 1;
Qy	1	CCCTGTATCTTTAAACCTCTTGTGTTAAAGTTTGTCTCTCCAGAATCAAAACTGTAAAACTA	60
Db	170822	CCCTGTATCTTTAAACCTCTTGTGTTAAAGTTTGTCTCTCCAAAATTGAAGCTCTAAACTA	170763
Qy	61	CAAAATGTTCTTCAAATGGAGACACAGATGGAGTCCATGACTAAGATCCACCGTGAGACC	120
Db	170762	CAAAATGTTCTTCAAATGGAGCCCCAGACGCGAGTCCATGACTAAGATCTACCGCAGACCC	170703
Qy	121	CTGGACCGGCGCTGAGCCCATGCTCCGATGTTAATGACATTGAAGGACACCCCTCCCGAG	180
Db	170702	CTGGACCGGCGCTGAGCCCATGCTCCGATGTTAATGACATCGAAGGACACCCCTCCAGAG	170643
Qy	181	GAATATCTCAACTGCACACACCCCTACTATGTCGCCCAATTTCAGCGGGAAGCAGTTAGAGCGGT	240
Db	170642	GAATATCTCAACTGCATTAACCCCTACTACGCCCAATTTCAGCAGGAAGCAGTTAGAGCACT	170583
Qy	241	CATCAGCCAACTCTCCCAACAGCACTTGGGTTTTCTGTTGAGAGGGGGACTGAGAGAC	300
Db	170582	CATCGGCCATCTCTCCCAACAGCACTTGGGTTTTCTGTTGAGAGGGGGTACTCAGAGAC	170523
Qy	301	AGGACTAGCTGGATTTCCTTAGGCCAAACGAAGATCCCTAAGCCTAGCTGGGAAGGTGACT	360
Db	170522	AGGACTAGCTGGATTTCCTTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT	170463
Qy	361	GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC-----	411
Db	170462	GCCTTACCTTTTAAACCCGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGGTAGGAA	170403
Qy	412	AGAGAGCTCACTAAATGCTTAATTAGGCAAAAAATAGGAGGTAAAGAAATAGCCAATCATC	471
Db	170402	AGAGAGCTCACTAAATGCTTAACCTAGGCTAAGGCTAAGGCTAAGGCTAAGGCTAAGGCT	170343
Qy	472	TATTGCGCTGAGACACAGCGGGGAGGACAGGATCGGGATATAAACCCAGGCAATTCGAGC	531
Db	170342	TATCGCGCTGAGAACACAGTGGGAGGGAACAATGTTGGGTATTAAACCCAGGCAATTCGAGC	170283
Qy	532	CGGCAACGGCAACCCCTTTTGGGTCCCTCCCTTTGTTATGGGCGCTCTGTTTCACTCTA	591
Db	170282	CAGCAATGGCTACCAATTTTGGGTCCCTCCCTTTGTTATGGGAGCTCTGTTTCACTCTA	170223
Qy	592	TTTCACCTCTATTAATCTTGGCACTGAAAAAATAAAAAAATAAAAAA	635
Db	170222	TTTCACCTCTATTAATCTTGGCACTGAAAAAATAAAAAAATAAAAAA	170179

RESULT 16


```

FT variation replace(53305,A)
FT /*tag= ar
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT variation replace(53378..53380,GG)
FT /*tag= as
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT exon 53378..53903
FT /*tag= at
FT 53904..83114
FT /*tag= au
FT variation replace(56233,G)
FT /*tag= av
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT variation replace(56453,C)
FT /*tag= aw
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT variation replace(59300,T)
FT /*tag= ax
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT variation replace(60998,A)
FT /*tag= ay
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT variation replace(63013,A)
FT /*tag= az
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT variation replace(63293,A)
FT /*tag= ba
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT variation replace(64181,T)
FT /*tag= bb
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT variation replace(64191,A)
FT /*tag= bc
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT variation replace(66992,C)
FT /*tag= bd
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT variation replace(76246,Y)
FT /*tag= be
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT variation replace(77368,G)
FT /*tag= bf
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT variation replace(78043,A)
FT /*tag= bg
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT variation replace(78395,T)
FT /*tag= bh
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT variation replace(79474,G)
FT /*tag= bi
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT variation replace(79992,T)
FT /*tag= bj
FT /standard_name= "Single nucleotide polymorphism (SNP)"

Query Match 83.9%; Score 532.8; DB 12; Length 326014;
Best Local Similarity 90.5%; Pred. No. 1.2e-153;
Matches 583; Conservative 0; Mismatches 52; Indels 9; Gaps 1;

QY 1 CCTGTATCTTTAACTCTCTGTTAAAGTTTGTCTCTTCCAGAAATCAAACTGTAAACTA 60
DB 170822 CCCTGTATCTTTAACTCTCTGTTAAAGTTTGTCTCTTCCAAATTTGAAGCTCTAAACTA 170763

QY 61 CAAATGTTCTTCAATGGAGCCAGCATGAGTCCATGATCAATGACATTCGAGCC 120
DB 170762 CAAATGTTCTTCAATGGAGCCAGCATGAGTCCATGATCAATGACATTCGAGCC 170703

QY 121 CTGACCGGCTGCTAGCCCATGCTCCGATGTTAAATGACATTCGAGCCATTCGAG 180
DB 170702 CTGACCGGCTGCTAGCCCATGCTCCGATGTTAAATGACATTCGAGCCATTCGAG 170643

QY 181 GAAATCTCACTGCAACCCCTACTATGCCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240

```

```

DB 170642 GAAATCTCACTGCAACCCCTACTAGCCCCCAATTCAGCGGGAAGCAGTTAGAGCAGT 170583
QY 241 CATCAGCCCACTCCCAACAGCACTTTGGGTTTTCCTGTTGAGAGGGGGGACTGAGAGAC 300
DB 170582 CATCGCCCATCTCCCAACAGCACTTTGGGTTTTCCTGTTGAGAGGGGGTACTCAGAGAC 170523
QY 301 AGGACTAGCTGGATTTCTTAGGCCAACGGAAGAAATCCCTAAGCCCTAGCTGGGAAAGTGACT 360
DB 170522 AGGACTAGCTGGATTTCTTAGGCCAACGGAAGAAATCCCTAAGCCCTAGCTGGGAAAGTGACT 170463
QY 361 GCATCCACCTCTAAACATGGGGCTTGCACCTTAGCTACACCCGACCAATC----- 411
DB 170462 GCTTCTACCTTTAAACCCGGGGCTTGCACCTTAGCTACACCCGACCAATCAGGTAGGAA 170403
QY 412 AGAGAGCTCACTAAAAATGCTAATTAGGCCAAAAATAGGAGTAAAGAAATAGCCAATCATC 471
DB 170402 AGAGAGCTCACTAAAAATGCTAATTAGGCCAAAAATAGGAGTAAAGAAATAGCCAATCATC 170343
QY 472 TATTGCTGAGAGCACAGCGGGAGGACAAAGGATCGGGATATTAACCCAGGCAATTCGAGC 531
DB 170342 TATCGGCTGAGAACACAGTGGGAGGACAATGATTGGGTTTATAAACCCAGGCAATTCGAGC 170283
QY 532 CGGCAAGCGCAACCCCTTTGGGTCCCTCCCTTGTATGGGCGCTCTGTTTCACTCTA 591
DB 170282 CAGCAATGGCTACCATTTTGGGTCCCTCCCTTGTATGGGAGCTCTGTTTCACTCTA 170223
QY 592 TTTCACCTATTAAATCTTGCAACTGAAAAAAGAAAAAAGAAAAA 635
DB 170222 TTTCACCTATTAAATCTTGCAACTGAAAAAAGAAAAAAGAAAAA 170179

RESULT 17
AAF55630
ID AAF55630 standard; DNA; 2781 BP.
XX AAF55630;
XX 29-MAY-2001 (first entry)
DE Nucleotide sequence of a human endogenous retrovirus envelope protein.
XX Envelope protein; HERV; syncytia formation; placental development;
KW syncytia; cancer; cell adhesion; ss.
XX Human endogenous retrovirus.
XX Key Location/Qualifiers
FT CDS 762..2378
FT /*tag= a
FT /product= "envelope protein"
XX WO200116171-A1.
XX 08-MAR-2001.
XX 01-SEP-2000; 2000WO-FR002429.
XX 01-SEP-1999; 99FR-00011141.
PR 15-SEP-1999; 99FR-00011793.
XX (INMR ) BIO MERIEUX.
PA (INRM ) INST NAT SANTE & RECH MEDICALE.
XX Mallet F, Cosset F, Blond J, Lavillette D, Bouton O, Ruggieri A;
XX WPI; 2001-226676/23.
DR P-PSDB; AAB67652.
XX Detecting expression of human endogenous retrovirus envelope protein in
PT cells of a tissue or culture, from its ability to induce syncytia.
XX Disclosure; Page 44-45; 57pp; French.
PS

```



```
PS Claim 1; Page 60-63; 106pp; French.
XX
XX This sequence represents clone cl.PH74 of the human endogenous retrovirus
CC (HERV) W genome. The nucleic acids, their fragments or peptides encoded
CC by them are markers of autoimmune disease (e.g. multiple sclerosis,
CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
CC dependent diabetes and related pathologies) and of abnormal or
CC unsuccessful pregnancy and can be used as chromosomal markers for
CC susceptibility to these conditions, or proximity markers of genes
CC associated with this susceptibility
XX
XX Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 U; 0 Other;
SQ
Query Match 82.3%; Score 522.6; DB 2; Length 2782;
Best Local Similarity 90.4%; Pred. No. 1.7e-151;
Matches 574; Conservative 0; Mismatches 49; Indels 12; Gaps 1;
QY 1 CCCTGTATCTTTAACTCTTGTAGTTGTTCTCTCCAGAAATCAAACTGTAAACTA 60
DB 2158 CCCTGTATCTTTGACCTCTTGTAACTTGTCTCTCCAGAAATCGAAGCTGTGA 2217
QY 61 CAAATGTTCTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGACCC 120
DB 2218 -----CAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGACCC 2265
QY 121 CTGGACCGGCTGTAGCCCATGCTCCGATGTTAATGACATTAAGAGCACCCTCCCGAG 180
DB 2266 CTGGACCGGCTGTAGCCCATGCTCCGATGTTAATGACATTAAGAGCACCCTCCCGAG 2325
QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAGCAGTTAGAGCGGT 240
DB 2326 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAGCAGTTAGAGCGGT 2385
QY 241 CATCAGCAACCTCCCAACAGCAGTGGTCTTCTGTTGAGAGGGGAGCTGAGAGAC 300
DB 2386 GGTGCGGCAACCTCCCAACAGCAGTGGTCTTCTGTTGAGAGGGGAGCTGAGAGAC 2445
QY 301 AGGACTAGCTGGATTTCTTAGGCCAAACGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
DB 2446 AGGACTAGCTGGATTTCTTAGGCCAAACGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 2505
QY 361 GCATCCACCTCTAACAATGCGGCTTGCACCTAGCTCAGACCCGACCAATCAGAGACTC 420
DB 2506 ACATCCACCTCTAACAATGCGGCTTGCACCTAGCTCAGACCCGACCAATCAGAGACTC 2565
QY 421 ACTAAATGCTAATAGGCAAAATAGGAGTAAAGAAATAGCAATCATCTATTGCTG 480
DB 2566 ACTAAATGCTAATAGGCAAAATAGGAGTAAAGAAATAGCAATCATCTATTGCTG 2625
QY 481 AGAGCACAGCGGAGGACAAAGGATCGGATATATAACCCAGGCAATTCGAGCCGCAACGG 540
DB 2626 AGAGCACAGCGGAGGACAAATGATCGGATATATAACCCAGGCAATTCGAGCCGCAACGG 2685
QY 541 CAACCCCTTTGGTCCCTCCCTTTGATGGGCTCTGTTTCACTCTATTCTACTCT 600
DB 2686 CAACCCCTTTGGTCCCTCCCTTTGATGGGCTCTGTTTCACTCTATTCTACTCT 2745
QY 601 ATTAAATCTTGCACTGAAAAAATAAAAAA 635
DB 2746 ATTAAATCTTGCACTGAAAAAATAAAAAA 2780
RESULT 21
AAAS9211
ID AAAS9211 standard; DNA; 2782 BP.
XX
XX AAAS9211;
AC
XX
XX
DT 07-NOV-2000 (first entry)
XX
XX 5' non coding, 3' pol gene, and 3' non coding sequences of HERV-W.
XX Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
KW
```

```
gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
XX Homo sapiens.
XX
XX WO2000043521-A2.
XX
XX 27-JUL-2000.
XX
XX 21-JAN-2000; 2000WO-FR000144.
XX
XX 21-JAN-1999; 99FR-00000888.
XX
XX (INMR ) BIO MERIEUX.
XX
XX Paranhos-Baccala G, Mallet F, Voisset C;
XX
XX WPI; 2000-499229/44.
XX
XX New nucleic acid from human endogenous retrovirus, useful e.g. for
XX diagnosis of autoimmune disease and complications of pregnancy, contains
XX at least part of the gag gene.
XX
XX Disclosure; Page 46-47; 53pp; French.
XX
XX The present sequence represents an endogenous retroviral nucleic acid
XX fragment, which is associated with an autoimmune disease, and is
XX integrated into the human genome. The fragment is originally derived from
XX a novel retrovirus, human endogenous retrovirus W (HERV-W). The HERV-W
XX retrovirus is associated with autoimmune disease, failure of pregnancy or
XX disorders of pregnancy. The nucleic acid fragment, or proteins derived
XX from it, are useful for diagnosis of autoimmune disease (specifically
XX multiple sclerosis) and for monitoring pregnancy. The nucleic acid
XX fragments may also be used for in situ labelling of isolated chromosomes,
XX while the transcription product can be used to study or monitor T cell
XX proliferation in vitro
XX
XX Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 U; 0 Other;
SQ
Query Match 82.3%; Score 522.6; DB 3; Length 2782;
Best Local Similarity 90.4%; Pred. No. 1.7e-151;
Matches 574; Conservative 0; Mismatches 49; Indels 12; Gaps 1;
QY 1 CCCTGTATCTTTAACTCTTGTAGTTGTTCTCTCCAGAAATCAAACTGTAAACTA 60
DB 2158 CCCTGTATCTTTGACCTCTTGTAACTTGTCTCTCCAGAAATCGAAGCTGTGA 2217
QY 61 CAAATGTTCTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGACCC 120
DB 2218 -----CAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGACCC 2265
QY 121 CTGGACCGGCTGTAGCCCATGCTCCGATGTTAATGACATTAAGAGCACCCTCCCGAG 180
DB 2266 CTGGACCGGCTGTAGCCCATGCTCCGATGTTAATGACATTAAGAGCACCCTCCCGAG 2325
QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAGCAGTTAGAGCGGT 240
DB 2326 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAGCAGTTAGAGCGGT 2385
QY 241 CATCAGCAACCTCCCAACAGCAGTGGTCTTCTGTTGAGAGGGGAGCTGAGAGAC 300
DB 2386 GGTGCGGCAACCTCCCAACAGCAGTGGTCTTCTGTTGAGAGGGGAGCTGAGAGAC 2445
QY 301 AGGACTAGCTGGATTTCTTAGGCCAAACGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
DB 2446 AGGACTAGCTGGATTTCTTAGGCCAAACGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 2505
QY 361 GCATCCACCTCTAACAATGCGGCTTGCACCTAGCTCAGACCCGACCAATCAGAGACTC 420
DB 2506 ACATCCACCTCTAACAATGCGGCTTGCACCTAGCTCAGACCCGACCAATCAGAGACTC 2565
QY 421 ACTAAATGCTAATAGGCAAAATAGGAGTAAAGAAATAGCAATCATCTATTGCTG 480
DB 2566 ACTAAATGCTAATAGGCAAAATAGGAGTAAAGAAATAGCAATCATCTATTGCTG 2625
```

QY 481 AGAGCACAGCGGAGGACAAAGATCGGATATTAACCCAGGACATTCAGCCGCGCAACGG 540
 |||||
 Db 2626 AGAGCACAGGAGGAGGACAAATGATCGGATATTAACCCAGGATTTTCGAGCCGCGCAACGG 2685
 |||||
 QY 541 CAACCCCTTTGGTCCCTCCCTTCTGATGGGCGCTCTGTTTCACTCTATTTCACCTCT 600
 |||||
 Db 2686 CAACCCCTTTGGTCCCTCCCTTCTGATGGGAGCTCTGTTTCACTCTATTTCACCTCT 2745
 |||||
 QY 601 ATTAATCTTGCAACTGAAAAAAGAAAAA 635
 |||||
 Db 2746 ATTAATCTTGCAACTGAAAAAAGAAAAA 2780
 |||||

RESULT 22

AAH20069

ID AAH20069 standard; DNA; 2782 BP.

XX AAH20069;

XX 08-AUG-2001 (first entry)

XX HERV-W envelope protein G encoding nucleic acid.

XX Human endogenous retrovirus; HERV-W; HERV; chromosome 7; env protein;
 KW envelope protein; multiple sclerosis-related superantigen; vaccine;
 KW surface antigen; transmembrane; multiple sclerosis; neuroprotective;
 KW antisense-therapy; autoimmune disorder; ds.

XX Human endogenous retrovirus.

XX Location/Qualifiers

FH Key 1..762

FT 5'UTR /*tag= a

FT CDS 763..2379

FT /*tag= b

FT /product= "HERV-W envelope protein G"

FT /transl_except= (pos:790..792,aa:Phe)

FT /transl_except= (pos:793..795,aa:Thr)

FT /transl_except= (pos:812..814,aa:Leu)

FT /transl_except= (pos:818..820,aa:Ser)

FT /transl_except= (pos:862..864,aa:Tyr)

FT /transl_except= (pos:865..867,aa:Gln)

FT /transl_except= (pos:1174..1176,aa:Arg)

FT /transl_except= (pos:1441..1443,aa:Leu)

FT /transl_except= (pos:1903..1905,aa:Gln)

FT /transl_except= (pos:2017..2019,aa:Iys)

FT /transl_except= (pos:2026..2028,aa:Arg)

FT /transl_except= (pos:2044..2046,aa:Arg)

FT /transl_except= (pos:2089..2091,aa:Gln)

FT /transl_except= (pos:2170..2172,aa:Asn)

FT 3'UTR 2380..2782

FT /*tag= c

XX WO200131021-A1.

XX 03-MAY-2001.

XX 30-OCT-2000; 2000WO-EP010659.

XX 28-OCT-1999; 99EP-00402690.

XX (UYGE-) UNIV GENEVE.

XX Conrad B, Mach B;

XX WPI; 2001-316336/33.

XX P-PSDB; AAB75138.

XX New human retrovirus HERV-W ENV proteins/peptides having superantigen
 PT activity useful for diagnosing and treating multiple sclerosis.
 XX Claim 13; Fig 9; 94pp; English.

PS

XX

CC On the basis of the PBS t-RNA motif used for the classification of human
 CC endogenous retrovirus (HERVs) the full length endogenous provirus which
 CC was located on the long arm of human chromosome 7 (7q21-22) has been
 CC designated HERV-W. The present invention describes proteins or peptides
 CC (I) having superantigen (SAG) activity comprising the ENV protein (ENV)
 CC of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)
 CC have neuroprotective activity, and can be used in: vaccines; antisense-
 CC therapy; and HERV-W SAG activity-inhibitors. (I) and encoding DNA/RNA are
 CC useful for diagnosing multiple sclerosis (MS) or HERV-W-associated
 CC disorders. (I) are also useful for identifying substances (and optionally
 CC recovering) capable of binding to a retroviral superantigen associated
 CC with MS, substances capable of blocking SAG activity and substances
 CC capable of blocking transcription or translation of HERV-W retroviral
 CC superantigen. A protein or peptide derived from (I), modified to be
 CC devoid of SAG activity and being capable of generating an immune response
 CC against HERV-W retroviral SAG is useful in therapy. Nucleic acid
 CC molecules encoding (I) are useful as vaccines against MS. Substances
 CC capable of blocking SAG activity, capable of binding to a retroviral
 CC superantigen associated with MS, or capable of blocking transcription or
 CC translation of HERV-W retroviral superantigen for use in treating or
 CC preventing MS, obtained using (I) are useful for the treatment and
 CC prevention of MS. (I) and nucleic acids encoding them are useful for
 CC diagnosing autoimmune disease. The present sequence encodes the
 CC specifically claimed envelope protein of HERV-W designated G

XX Sequence 2782 BP; 741 A; 768 C; 564 G; 709 T; 0 U; 0 Other;

SQ Query Match 82.3%; Score 522.6; DB 5; Length 2782;

Best Local Similarity 90.4%; Pred. No. 1.7e-151;

Matches 574; Conservative 0; Mismatches 49; Indels 12; Gaps 1;

QY 1 CCCTGTATCTTTAAACCTCCTTGTAAAGTTTGTCTTCTCCAGAAATCAAACTGTAAACTA 60

Db 2158 CCCTGTATCTTTGACCTCCTTGTAACTTGTCTTCTCCAGATCGAGCTGTGAACCTA 2217

QY 61 CAAATTGTTCTTCAAAATGAGACACAGATGGAGTCCATGACTAAGATCCACCGTGACCC 120

Db 2218 -----CAAAATGGAGGCCAAAGATGCAGTCCCAAGACTAAGATCTACCGCAGACCC 2265

QY 121 CTGGACCGGCTGTAGCCCATGCTCCGATGTTAATGACATTCAGAGCACCCCTCCCGAG 180

Db 2266 CTGGACCGGCTGTAGCCCATGCTCCGATGTTAATGACATTCAGAGCACCCCTCCCTGAG 2325

QY 181 GAAATCTCAATGTCACAAACCCCTACTATGCCCCAAATTCAGCGGGAAGACAGTTCAGCGGT 240

Db 2326 GAAATCTCAATGTCACAAACCCCTACTATGCCCCAAATTCAGCGGGAAGACAGTTCAGCGGT 2385

QY 241 CATCAGCCAACTCCCAACAGACATTTGGGTTTTCTGTTGAGAGGGGGAGCTGAGAGAC 300

Db 2386 GGTGGGCCAACTCCCAACAGACATTTAGGTTTTCTGTTGAGATGGGGAGCTGAGAGAC 2445

QY 301 AGGACTAGCTGGATTTTCTTAGGCCAACGAGATCCCTAAGCCTAGCTGGGAGGTGACT 360

Db 2446 AGGACTAGCTGGATTTTCTTAGGCCAACGAGATCCCTAAGCCTAGCTGGGAGGTGACT 2505

QY 361 GCATCCACCTTAAACATGGGCTTGCAACTTAGCTCACACCCGACCAATTCAGAGAGCTC 420

Db 2506 ACATCCACCTTAAACATGGGCTTGCAACTTAGCTCACACCCGACCAATTCAGAGAGCTC 2565

QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCCAAATCATCTATTGCGCTG 480

Db 2566 ACTAAATGCTAATTAGGCAAAAGAGGAGGTAAAGAAATAGCCAAATCATCTATTGCGCTG 2625

QY 481 AGAGCACAGGAGGAGGACAGATCGGATATTAACCCAGGACATTCGAGCCGCGCAACGG 540

Db 2626 AGAGCACAGGAGGAGGACAGATCGGATATTAACCCAGGACATTCGAGCCGCGCAACGG 2685

QY 541 CAACCCCTTTGGGTCCTCCCTTCTGTTGAGGCGCTCTGTTTCACTCTATTTCACCTCT 600

Db 2686 CAACCCCTTTGGGTCCTCCCTTCTGTTGAGGAGCTCTGTTTCACTCTATTTCACCTCT 2745

QY 601 ATTAATCTTGCAACTGAAAAAAGAAAAA 635

||||| 2746 ATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAAAAAA 2780

Db 2433 CTGGACCGGCTGCTAGCCACGATCTGATTTAATGACATCAAGGCGACCCCTCTCTGAG 2492

QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAAAGCAGTTAGAGCGGT 240

Db 2493 GAAATCTCAGCTGCACAACTCTACTACGCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2552

QY 241 CATAGCCAACTCCCAACAGACCTTGGGTTTTCTGTGTGAGAGGGGGAGCTGAGAGAC 300

Db 2553 CGTGGGCCAACTCCCAACAGACCTTAGGTTTTCTGTGTGAGATGGGGAGCTGAGAGAC 2612

QY 301 AGGACTAGCTGGATTTCTTAGGCCAACGAAATCCCTAAGCCTAGCTGGAAGGTGACT 360

Db 2613 AGGACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGGAAGGTGACT 2672

QY 361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACAACCCGACCAATCAGAGAGCTC 420

Db 2673 ACATCCACCTTTAAACACAGGGGCTTGCAACTTAGCTCACAACCTGACCAATCAGAGAGCTC 2732

QY 421 ACTAAATGCTAATTAGGCAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTCGCTG 480

Db 2733 ACTAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTCGCTG 2792

QY 481 AGAGCACGGGGAGGAGCAAGATCGGATATACCCAGGCAATTCGAGCGGCAACGG 540

Db 2793 AGAGCACAGCAGGAGGAGCAATGATCGGATATACCCAGGCAATTCGAGCGGCAACGG 2852

QY 541 CAACCCCTTTGGGTCCTCCCTTGTGTATGGCGCTCTGTTTTCATCTCTATTTCACTCT 600

Db 2853 CAACCCCTTTGGGTCCTCCCTTGTGTATGGCGCTCTGTTTTCATCTCTATTTCACTCT 2912

QY 601 ATTAAATCTTGCAACTG 617

Db 2913 ATTAAATCTTGCAACTG 2929

RESULT 24

ABL61744

ID ABL61744 standard; DNA; 56093 BP.

XX AC ABL61744;

XX DT 15-MAY-2002 (first entry)

XX DE Colon adenocarcinoma related gene sequence SEQ ID NO:81.

XX DE Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;

XX KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;

XX KW cytoskeletal; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;

XX KW gene; ds.

XX OS Homo sapiens.

XX PN WO200194629-A2.

XX PD 13-DEC-2001.

XX PF 30-MAY-2001; 2001WO-US010838.

XX PR 05-JUN-2000; 2000US-0209473P.

XX PR 05-JUN-2000; 2000US-0209531P.

XX PR 18-SEP-2000; 2000US-0233133P.

XX PR 18-SEP-2000; 2000US-0233617P.

XX PR 20-SEP-2000; 2000US-0234009P.

XX PR 20-SEP-2000; 2000US-0234034P.

XX PR 20-SEP-2000; 2000US-0234052P.

XX PR 22-SEP-2000; 2000US-0234509P.

XX PR 22-SEP-2000; 2000US-0234567P.

XX PR 25-SEP-2000; 2000US-0234923P.

XX PR 25-SEP-2000; 2000US-0234924P.

XX PR 25-SEP-2000; 2000US-0235077P.

XX PR 25-SEP-2000; 2000US-0235082P.

XX PR 25-SEP-2000; 2000US-0235134P.

Db 2746 ATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAAAAAA 2780

RESULT 23

AAD24195

ID AAD24195 standard; cDNA; 2930 BP.

XX AC AAD24195;

XX DT 07-MAY-2002 (first entry)

XX DE Human syncytin cDNA.

XX KW Human; syncytin; preeclampsia; gestational trophoblast disorder;

XX KW choriocarcinoma; hydatiform mole; placental site tumour; abortion;

XX KW envelope gene; human endogenous defective retrovirus; HERV-W; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 930..2546

XX FT /*tag= a

XX FT /product= "Syncytin"

XX PN WO200204678-A2.

XX PD 17-JAN-2002.

XX PF 09-JUL-2001; 2001WO-US021719.

XX PR 07-JUL-2000; 2000US-0216657P.

XX PA (GEMV) GENETICS INST INC.

XX PI Keith JC, McCoy JM, Mi S;

XX WPI; 2002-171727/22.

XX P-PSDB; AAE14540.

XX PT Identifying a compound for treating a subject with or at risk of

XX PT developing preeclampsia, comprises determining whether the expression or

XX PT activity of syncytin in the cell is modulated in the presence of a test

XX PT compound.

XX PS Disclosure; Page 39-42; 43pp; English.

XX CC The invention relates to identifying compounds which are modulators of

XX CC syncytin expression. The syncytin modulators are useful in diagnosis and

XX CC treatment of preeclampsia and gestational trophoblast disorders (e.g.

XX CC choriocarcinoma, hydatiform mole, placental site tumour and missed/

XX CC incomplete abortion). Syncytin is a human gene derived from the envelope

XX CC gene of human endogenous defective retrovirus, HERV-W. The present

XX CC invention is based partly on the discovery that syncytin expression is

XX CC dramatically reduced in preeclampsia, and is also mis-localised to the

XX CC apical syncytiotrophoblast membrane. The present sequence is human

XX CC syncytin cDNA

XX SQ Sequence 2930 BP; 842 A; 800 C; 571 G; 717 T; 0 U; 0 Other;

Query Match 81.5%; Score 517.4; DB 6; Length 2930;

Best Local Similarity 91.4%; Pred. No. 7.4e-150;

Matches 564; Conservative 0; Mismatches 41; Indels 12; Gaps 1;

QY 1 CCCTGTATCTTTAACTCTCTGTTAGTTTGTCTCTCCAGATCAAACTGTAATACTA 60

Db 2325 CCCTGTATCTTTAACTCTCTGTTAGTTTGTCTCTCCAGATCGAAGCTGTAATACTA 2384

QY 61 CAAATTTGTTCTCAATGGAGCACCAGATGAGTCCATGATCAAGATCCACCTGGACCC 120

Db 2385 -----CAAATGGAGCCCAAGATGAGTCCAGATCAAGATCTACCGCAGACC 2432

QY 121 CTGGACCGGCTGCTAGCCCATGCTCCGATGTTAATGACATTTGAAGCGACCCCTCCCGAG 180

```
PR 25-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 28-SEP-2000; 2000US-0236442P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237284P.
PR 02-OCT-2000; 2000US-0237295P.
PR 02-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX
PA (AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX
XX Claim 1; SEQ ID NO 81; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC tumour
XX
SQ Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16881 T; 0 U; 0 Other;
Query Match 81.3%; Score 516.2; DB 6; Length 56093;
Best Local Similarity 91.1%; Pred. No. 7.4e-149;
Matches 564; Conservative 0; Mismatches 43; Indels 12; Gaps 1;
QY 1 CCTGTATCTTTAACTCTTGTAGTTGTCTCTTCCAGATCAAACTGTAAACTA 60
DB 37274 CCTGTATCTTTAACTCTTGTAGTTGTCTCTTCCAGATCGAAGCTGTAAACTA 37333
QY 6i CAAATGTTCTTCAATGGAGCACCAGATCGAGTCCATGACTAAGATCCACCGTGACCC 120
DB 37334 -----CAAATGGAGCACCAGATCGAGTCCAGACTAAGATCTACCGCAGACCC 37381
```

```
QY 121 CTGGACCGGCTGTAGCCCATGCTCCGATGTTTAATGACATTTGAAGCACCCTCCCGAG 180
DB 37382 CTGGACCGGCTGTAGCCCATGCTCCGATGTTTAATGACATCAAAGGCACCCCTCTCTGAG 37441
QY 181 GAAATCTCACTGACACAAACCCCTACTATGCCCCCAATTCAGGGGAAGCAGTTAGACGGT 240
DB 37442 GAAATCTCACTGACACAAACCCCTACTATGCCCCCAATTCAGCGGAAGCAGTTAGACGGT 37501
QY 241 CATCAGCCAACTCCCCCAACAGCACTTTGGGTTTTTCTGTGTTGAGAGGGGGGACTGAGAGAC 300
DB 37502 CGTCGGCCAACTCCCCCAACAGCACTTAGGTTTTTCTGTGTTGAGATGGGGACTGAGAGAC 37561
QY 301 AGGACTAGCTGGAATTTCTTAGGCCAACAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
DB 37562 AGGACTAGCTGGAATTTCTTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGGGAAGGTGACC 37621
QY 361 GCATCCACCTCTAAACATGCGGGCTTGCAACTTAGCTACACCCGACCAATCAGAGAGCTC 420
DB 37622 ACATCCACCTTTAAACACGCGGGCTTGCAACTTAGCTACACCCGACCAATCAGAGAGCTC 37681
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCCAAATCATCTATTGCGCTG 480
DB 37682 ACTAAATGCTAATTAGGCAAAAGACAGCAGGAGTAAAGAAATAGCCAAATCATCTATTGCGCTG 37741
QY 481 AGAGCAGCGGAGGAGCAAGGATCGGGATATAAACCAGGCAATTCGAGCCGGCAACGG 540
DB 37742 AGAGCAGCAGCAGGAGGACAAATGATCGGGATATAAACCAGGCAATTCGAGCCGGCAACGG 37801
QY 541 CAACCCCTTTGGGTCCTCCCTTTGTATGGGGCTCTGTTTTCATCTATTTCATCTCT 600
DB 37802 CAACCCCTTTGGGTCCTCCCTTTGTATGGGAGCTCTGTTTTCATCTATTTCATCTCT 37861
QY 601 ATTAATCTTGCAACTGAA 619
DB 37862 ATTAATCTTGCAACTGCA 37880
RESULT 25
AAK25665
ID AAK25665 standard; cDNA to mRNA; 7582 BP.
XX
AC AAK25665;
XX
DT 21-MAY-1999 (first entry)
XX
DE Complete human endogenous retrovirus W genome.
XX
KW Clone; human endogenous retrovirus; genome; autoimmune disease;
KW multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KW disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
XX
OS Human endogenous retrovirus.
XX
XX WO9902696-A1.
XX
PD 21-JAN-1999.
XX
PF 06-JUL-1998; 98WO-FR001442.
XX
PR 07-JUL-1997; 97FR-00008815.
XX
XX (INMR ) BIO MERIEUX.
XX
PI Besene F, Blond J, Bouton O, Mandrand B, Mallet F;
XX
XX WPI; 1999-120897/10.
XX
XX New nucleic acid sequences from human endogenous retrovirus-W - expressed
PT exclusively in placenta and useful in diagnosis and therapy of autoimmune
PT disease, and abnormal or failed pregnancy.
XX
XX Claim 1; Page 71-74; 106pp; French.
PS
```


Db 7036 -----CAAATGGAGCCCAAGATGCAGTCCAAAGACTAAGATCTACCGAGACC 7083

Qy 121 CTGACCGGCTGTAGCCCATGCTCCGATGTTAATGACATTAAGGACACCCCTCCCGAG 180

Db 7084 CTGACCGGCTGTAGCCCATGCTCCGATGTTAATGACATTAAGGACACCCCTCCCGAG 7143

Qy 181 GAAATCTCAATGACACACCCCTACTATGCCCCAAATTCAGCGGAGCAGTTAGAGGCT 240

Db 7144 GAAATCTCAATGACACACCCCTACTATGCCCCAAATTCAGCGGAGCAGTTAGAGGCT 7203

Qy 241 CATCAGCAACCTCCCAACAGCACCTTGGGTTTCTGTTGAGAGGGGAGCTGAGAGAC 300

Db 7204 SGTGCGCAACCTCCCAACAGCACCTTGGGTTTCTGTTGAGATGGGAGCTGAGAGAC 7263

Qy 301 AGACTAGCTGATTTCTTAGGCAACGAAAGATCCCTAAGCCTAGCTGGGAAGGTGACT 360

Db 7264 AGACTAGCTGATTTCTTAGGCTGATTAAGAAATCCCTAAGCCTAGSTGGGAAGGTGACC 7323

Qy 361 GCATCCACCTTAACATGGGCTTGCACCTAGCTCAGACCCGACCAATCAGAGAGCTC 420

Db 7324 ACATCCACCTTAAACACGGGGCTTGCACCTAGCTCAGACCCGACCAATCAGAGAGCTC 7383

Qy 421 ACTAAATGCTTAATTAGGCAAAATAGCAGTAAAGAAATAGCAATCATCTATTGCTG 480

Db 7384 ACTAAATGCTTAATTAGGCAAGACAGAGGTAAGAAATAGCAATCATCTATTGCTG 7443

Qy 481 AGAGCAGCGGAGGAGCAAGGATCGGGATATAAACCAGGATTCGAGCCGCAACGG 540

Db 7444 AGAGCAGCAGGAGGAGCAATGATCGGATATAAACCAGGATTCGAGCCGCAACGG 7503

Qy 541 CAACCCCTTTGGGTCCTCCCTTGTATGGGGCTCTGTTTCACTCTATTTCATCT 600

Db 7504 CAACCCCTTTGGGTCCTCCCTTGTATGGGGCTCTGTTTCACTCTATTTCATCT 7563

Qy 601 ATTAAATCTTGCAACTG 617

Db 7564 ATTAAATCTTGCAACTG 7580

RESULT 27
ABN97978

ID ABN97978 standard; DNA; 46340 BP.

XX AC ABN97978;

XX DT 01-AUG-2002 (first entry)

XX DE Human retroviral sequence H13.

XX KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
multiple sclerosis; ds.

XX OS Human endogenous retrovirus.

XX PN WO967395-A1.

XX PD 29-DEC-1999.

XX PF 23-JUN-1999; 99WO-FR001513.

XX PR 23-JUN-1998; 98FR-00007920.

XX XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX PA Alliel PM, Perin J, Rieger F;

XX XX WPI; 2000-160587/14.

XX PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
for diagnosis, treatment and prevention of autoimmune and neurological
diseases.

XX PS Claim 15; Page 186-199; 225pp; French.

XX

CC The present invention relates to new nucleic acid sequences of human
endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention

XX SQ Sequence 46340 BP; 16104 A; 8738 C; 8434 G; 13064 T; 0 U; 0 Other;

Query Match 80.4%; Score 510.6; DB 3; Length 46340;

Best Local Similarity 90.0%; Pred. No. 3.7e-147;

Matches 575; Conservative 0; Mismatches 49; Indels 15; Gaps 2;

Qy 1 CCCTGTATCTTTAACTCTCTTGTAAAGTTTGTCTCTTCCAGAAATCAAACTGTAATACTA 60

Db 33280 CCCTGTATCTTTAACTCTCTTGTAAAGTTTGTCTCTTCCAGAAATCAAACTGTAATACTA 33339

Qy 61 CAAATTTGTTCTTCAAATGGAGCACAGATGGAGTCCATGATCAAGATCCACCGTGGACCC 120

Db 33340 CAAATCGTTCTTCAAATGGAAACCCAGATGAAGTCCATGATCAAGATCCACCGTGGACCC 33399

Qy 121 CTGACCGGCTCTAGCCCATGCTCCGATGTTAATGACATTAAGGACACCCCTCCCGAG 180

Db 33400 CTGACCGGCTCTAGCCCATGCTCCGATGTTAATGACATTAAGGACACCCCTCCCGAG 33459

Qy 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCCAAATTCAGCGGAGCAGTTAGAGCGGT 240

Db 33460 GAAATCTCAACTGCACAAACCCCTACTATGCCCCAAATTCAGCGGAGCAGTTAGAGCGGT 33519

Qy 241 CATCAGCCAACTCCCAACAGCACCTTGGGTTTCTGTTGAGAGGGGGAGCTGAGAGAC 300

Db 33520 CGTCAGCCAACTCCCAACAGCACCTTGGGTTTCTGTTGAGAGGGGGAGCTGAGAGAC 33579

Qy 301 AGGACTAGCTGGATTTCTTAGGCCAACGACAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACT 360

Db 33580 AGGATTAGCTGGATTTCTTAGGCCGACTAAGAAATCCCAAGCCCTAGCTGGGAAGGTGACC 33639

Qy 361 GCATCCACCTTAACATGCGGCTTGCAACTAGCTCAGACCCGACCAATC----- 411

Db 33640 ACATCCACCTTTAAACACTGGGCTTGCAACTAGCTCAGACCCGACCAATCAGGTAGTAA 33699

Qy 412 AGAGAGCTCACTAAATGCTTAATTAGGCAAAATAGGAGGTAAAGAAATAGCCAATCATC 471

Db 33700 AGAGAGCTCACTAAATGCTTAATTAGCAAAAACAGGAGTAAATAATAGCCAATCATC 33759

Qy 472 TATTGCTGAGAGCACAGCGGGAGGACAAAGATTCGGGATATAAACCCAGGCAATTCGAGC 531

Db 33760 TATCGCTGAGAGCACAGCGGGAGGACAAATGATCGGGATATAAACCCAGGCAATTCGAGC 33819

Qy 532 CGGCAACGGCAACCCCTTTGGGTCCTCCCTTTCTGTTGAGGGGCTCTGTTTCACTCTA 591

Db 33820 CGGCAACGGCTACCTTTCTTTGGGTCCTCCCTTTCTGTTGAGGGAGCTCTCTCT-----GT 33873

Qy 592 TTTCACCTATTAAATCTTGCAACTGAAAAAATAATA 630

Db 33874 TTTCACCTATTAAATCTTGCAACTGAAAAAATAATA 33912

RESULT 28

ABN97929

ID ABN97929 standard; DNA; 10499 BP.

XX AC ABN97929;

XX DT 01-AUG-2002 (first entry)

XX Human retroviral sequence HERV-7q.
DE Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
XX multiple sclerosis; ds.
KW Human endogenous retrovirus.
XX WO9967395-A1.
XX 29-DEC-1999.
XX 23-JUN-1999; 99WO-FR001513.
XX 23-JUN-1998; 98PR-00007920.
XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX Alliel PM, Perin J, Rieger F;
XX WPI; 2000-160587/14.
XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
PT for diagnosis, treatment and prevention of autoimmune and neurological
PT diseases.
XX Claim 3; Fig 1; 225pp; French.
XX The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention
XX
SQ Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 U; 0 Other;
Query Match 79.7%; Score 505.8; DB 3; Length 10499;
Best Local Similarity 91.1%; Pred. No. 5.6e-146;
Matches 564; Conservative 0; Mismatches 42; Indels 13; Gaps 2;
1 CCTGTATCTTTAACTCTCTTTAAAGTTTGTCTCTTCCAGATCAAACTGTAAACTA 60
9274 CCTGTATCTTTAACTCTCTTTAAAGTTTGTCTCTTCCAGATCAAACTGTAAACTA 9333
61 CAAATGTTCTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCAGCGTGACCC 120
9334 -----CAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCAGCGTGACCC 9381
121 CTGACCGGCGTCTGACCGCATGCTCCGATGTTAATGACATTCAGAGCACCCTCCCGAG 180
9382 CTGACCGGCGTCTGACCGCATGCTGATGTTAATGACATTCAGAGCACCCTCCCTGAG 9441
181 GAAATCTCAATGACAAACCCCTACTATGCCCCCAATTCAGCGGAGACAGTTAGACGGT 240
9442 GAAATCTCAATGACAAACCCCTACTATGCCCCCAATTCAGCGGAGACAGTTAGACGGT 9501
241 CATGACCAACCTCCCAACAGACTGGGTTTCTTGTGAGAGGGGAGCTGAGAGAC 300
9502 C-TGGGCAACCTCCCAACAGACTGGGTTTCTTGTGAGATGGGGAGCTGAGAGAC 9560
301 AGACTAGCTGGATTTCTTAGGCCAAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACT 360
9561 AGACTAGCTGGATTTCTTAGGCCAAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACT 9620
361 GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420

Db 9621 ACATCCACCTTTAAACACGGGCTTGCACCTTAGCTCACCTTGACCAATCAGAGAGCTC 9680
Qy 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCCAAATCATCTATTGCTTG 480
Db 9681 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCCAAATCATCTATTGCTTG 9740
Qy 481 AGAGCACAGCGGAGGACCAAGGATCGGATATATAACCCAGGATTCGAGCCGCAACGG 540
Db 9741 AGAGCACAGCGGAGGACCAAGGATCGGATATATAACCCAGGATTCGAGCCGCAACGG 9800
Qy 541 CAACCCCTTTGGTCCCTCCCTTTGTTATGGGCGCTCTGTTTCACTCTATTCTACTCT 600
Db 9801 CAACCCCTTTGGTCCCTCCCTTTGTTATGGGCGCTCTGTTTCACTCTATTCTACTCT 9860
Qy 601 ATTAAATCTTGCAACTGAA 619
Db 9861 ATTAAATCTTGCAACTGCA 9879
RESULT 29
AA114608
ID AA114608 standard; DNA; 1894 BP.
XX AC AA114608;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #4541 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000670.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human cervical epithelial cells.
XX
XX Claim 25; SEQ ID NO 4541; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging of
CC diseases of the cervix, notably cervical cancer. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;

Query Match	78.7%	Score 499.6	DB 4	Length 1894
Best Local Similarity	88.7%	Pred. No. 2.1e-144		
Matches 571	Conservative 0	Mismatches 54	Indels 19	Gaps 2
Qy 1	CCCTGTATCTTTAAACCTCTTGTAAAGTTTGTCTCTCCAGAAATCAAAATGTAATAACTA	60		
Db 1217	CCCTGTATCTTTAAACCTCTTGTAAAGTTTGTCTCTCCAGAAATCAAAATGTAATAACTA	1276		
Qy 61	CAAAATTTGTTCTTCAAAATGGAGCACACAGATGGAGTCCATGACTAAGATCCACCGTGGACCC	120		
Db 1277	CAAAATCGTTCTTCAAAATGGAGCCCACAGATGCAGTCCATGAGTAAATCTACCAGGACCC	1336		
Qy 121	CTGACCGGCTCTGCTAGCCCATGCTCCGATGTTTAAATGACATTTGAAGGACACCCCTCCCGAG	180		
Db 1337	CTGACCGGCTCTGCTAGCCCATGCTCTGATGTTTAAATGACATCAAGGACACCCCTCCCGAG	1396		
Qy 181	GAATATCTCAACTGACACACCCCTACTATGCCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT	240		
Db 1397	GAATATCTCAACTGACACAACTCTACTAGCCCCCAATTCAGCAGGAAGCAGTTAGAGTGGT	1456		
Qy 241	CATCAGGCCAACCTCCCAACAGCACACTTGGGTTTTCTGTTGAGAGGGGGGACTGACAGAC	300		
Db 1457	TGTTGGCCAACTCCCAACAGCACAGTTGGGTTTTCTGTTGAGAGGGGGGACTGACAGAC	1516		
Qy 301	AGGACTAGCTGGATTTCTTAGGCCAACGAAAGATCCCTTAAGCTTAGCTGGGAAGGTGACT	360		
Db 1517	AGGAAATAACTAGATTCTCTAGACCAACTAAGAAATCCCTAAGACTTAGCTGGGAAGGTGACC	1576		
Qy 361	GCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC-----	411		
Db 1577	GCTTCCACCTTTAAACACCGGGCTTGCAACTTAGCTCACACCCGACCAATC-----	411		
Qy 412	AGAGAGCTCACTAAATGCTAAATTAGGCAAAAAATAGGAGTTAAAGAAATAGCCAAATCATC	471		
Db 1637	AGAGAGCTCACTAAATGCTAAATTAGGCAAAAAACAGGAGATTAAGAAATAGCCAAATCATC	1696		
Qy 472	TATTGCTGAGAGCACACGGGAGGGAACAGGATCGGGATATAAACCCAGGCATTCGAGC	531		
Db 1697	TGTTGCCCTGACAGCACAGCAGGAGGGAACAATGATCGGGATATAAACCCAGGCATTCGAGC	1756		
Qy 532	CGGCAACGGCAACCCCTTTGGGTCCCTCTGTTGATGGCGCTCTGTTTCACTCTA	591		
Db 1757	CAGCTACAGCTACCCTCTTTGGGTCCCTCTGTTGATGGCGCTCTGTTTCACTCTA	1806		
Qy 592	TTTCACCTCTATTAAATCTTGCAACTGAAAAAATAAAAAAATAAAAAA	635		
Db 1807	CTTCACCTCTATTAAATCTTGCAACTGAAAAAATAAAAAAATAAAAAA	1850		
RESULT 30				
ABAS56337				
ID	ABAS56337 standard; DNA; 1894 BP.			
XX	AC	ABAS56337;		
XX	AC	ABAS56337;		
XX	DT	01-FEB-2002 (first entry)		
XX	DE	Human foetal liver single exon nucleic acid probe #4642.		
XX	DE	Human, foetal liver; gene expression; single exon nucleic acid probe; ss.		
XX	OS	Homo sapiens.		
XX	XX	W0200157277-A2.		
XX	PD	09-AUG-2001.		
XX	PF	30-JAN-2001; 2001WO-US000669.		
XX	XX	04-FEB-2000; 2000US-0180312P.		
XX	PR	26-MAY-2000; 2000US-0207456P.		
XX	PR	30-JUN-2000; 2000US-0608408.		
XX	PR	03-AUG-2000; 2000US-00632366.		

PR	21-SEP-2000; 2000US-0234687P.	
PR	27-SEP-2000; 2000US-0236359P.	
PR	04-OCT-2000; 2000GB-00024263.	
XX		
PA	(MOLE-) MOLECULAR DYNAMICS INC.	
XX		
PI	Penn SG, Hanzel DK, Chen W, Rank DR;	
XX		
DR	WPI; 2001-483447/52.	
XX		
PT	Human genome-derived single exon nucleic acid probes useful for analyzing	
PT	gene expression in human fetal liver.	
XX		
XX	Claim 1; SEQ ID NO 4642; 639pp + Sequence Listing; English.	
XX		
CC	The invention relates to a single exon nucleic acid probe for measuring	
CC	human gene expression in a sample derived from human foetal liver. The	
CC	single exon nucleic acid probes may be used for predicting, measuring and	
CC	displaying gene expression in samples derived from human fetal liver. The	
CC	present sequence is a single exon nucleic acid probe of the invention. The	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published_pct_sequences	
XX		
SQ	Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;	
	Query Match 78.7%; Score 499.6; DB 4; Length 1894;	
	Best Local Similarity 88.7%; Pred. No. 2.1e-144;	
	Matches 571; Conservative 0; Mismatches 54; Indels 19; Gaps 2;	
Qy	1 CCCTGTATCTTTAACTCTCTTGTTAAAGTTTGTCTTTCCAGAATCAAAACTGTAAAACTA 60	
Db	1217 CCCTGTATCTTTAACTCTCTTGTTAAAGTTTGTCTTTCCAGAATCGAAGCAGTAAAACTA 1276	
Qy	61 CAAATTTCTTCCAATGGAGCACCAGATCGAGTCCATGACTCAAGATCCACCGTGGACCC 120	
Db	1277 CAAATGCTTTCCAATGGAGCACCAGATCGAGTCCATGAGTAAATCTACCAAGGACCC 1336	
Qy	121 CTGACCGGGCTGCTAGCCCATGCTCCGATGTTTAATGACATTTGAAGCACCCTCCCGAG 180	
Db	1337 CTGACCGGGCTGCTAGCCCATGCTCTGATGTTAATGACATCAAGCACCCTCCCGAG 1396	
Qy	181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAAGCAGTTAGAGCGGT 240	
Db	1397 GAAATCTCAACTGCACAAACCTTACTACGCCCAATTCAGCAGGAAGCAGTTAGAGTGGT 1456	
Qy	241 CATCAGCAACCTCCCAACAGACACTTGGGTTTTCTGTTGAGAGGGGAGCACTGAGAGAC 300	
Db	1457 TGTGTGGCAACCTCCCAACAGCAGTTGGGTTTTCTGTTGAGAGGGGAGCACTGAGAGAC 1516	
Qy	301 AGGACTAGCTGGATTCTTCTTAGGCCAACGAAGAATCCCTAAGCCTAGCTGCGGAAGGTGACT 360	
Db	1517 AGGATAACTAGATTCTTAGACCAACTAGAAATCCCTAGACATAGCTGCGGAAGGTGACC 1576	
Qy	361 GCATCCACCTCTAAACATATGGGGCTTGCAACTTAGCTCACACCCGACCAATC----- 411	
Db	1577 GCTTCCACCTTTAAACACACGGGGCTTGCAACTTAGCTCACACCCCAACCAATCAGATACTAA 1636	
Qy	412 AGAGAGCTCACTAAATGCTAATTAGGCAAAAATAGGAGTAAGAATAAGCAATCATC 471	
Db	1637 AGAGAGCTCACTAAATGCTAATTAGGCAAAAACAGGAGATAAGAATAAGCAATCATC 1696	
Qy	472 TATTGCTGAGAGCACAGCGGGAGGCAAGGATCGGGATATAAACCCAGGCATTCGAGC 531	
Db	1697 TGTGTGCTGACAGCACAGGAGGAGCAATGATCGGGATATAAACCCAGGCATTCGAGC 1756	
Qy	532 CGGCAACGGCAACCCCTTTTGGGTCCCTCTCTTTGTATGGGGCTCTGTTTTCTACTCTA 591	
Db	1757 CAGCTACAGCTACCCCTTTTGGGTCCCTCTCTTTGTATGGGAGCTCTGT----- 1806	
Qy	592 TTTTCACTCTATTAATCTTCGAACTGAAAAAATAAAAAA 635	
Db	1807 CTTTCACTCTATTAATCTTCGAACTGAAAAAATAAAAAA 1850	

Search completed: January 22, 2005, 16:40:17
Job time : 302.418 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2005, 12:03:14 ; Search time 58.7997 Seconds
(without alignments)
7676.074 Million cell updates/sec

Title: US-09-319-156B-6
Perfect score: 635
Sequence: 1 cccgtatctttaacctctc.....tgaaaaaaaaaaaaaaaaaaaaa 635

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A-COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B-COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A-COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B-COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS-COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	635	100.0	635	4	US-08-979-847B-102
2	545	85.8	1329	4	US-08-979-847B-108
3	533.8	84.1	2946	3	US-09-175-928-3
4	233.6	36.8	279	1	US-08-686-878A-50
5	233.6	36.8	279	1	US-08-721-489-4
6	196.6	31.0	456	4	US-09-621-976-9366
7	128	20.2	1859	3	US-08-691-563C-46
8	128	20.2	1859	4	US-09-374-766-46
9	128	20.2	1859	4	US-08-979-847B-42
10	125.8	19.8	80246	3	US-09-078-294-4
11	125.8	19.8	80595	3	US-09-078-294-3
12	125	19.7	1722	3	US-08-691-563C-58
13	125	19.7	1722	4	US-09-374-766-58
14	125	19.7	1722	4	US-08-979-847B-54
15	110.6	17.4	149	4	US-09-513-999C-24228
16	76.4	12.0	3910	3	US-09-120-653D-1
17	73.4	11.6	492	4	US-09-621-976-11057
18	71.8	11.3	443	4	US-09-513-999C-21394
19	68.4	10.8	1481	4	US-08-979-847B-105
20	41.4	6.5	342	4	US-09-513-999C-13878
21	41.2	6.5	3824	4	US-10-140-002-541
22	40.4	6.4	464	3	US-09-120-653D-6
23	38.6	6.1	47	4	US-09-422-978-3579
24	37.6	5.9	7218	1	US-08-232-463-14
25	36	5.7	396	4	US-09-854-133-539
26	35.2	5.5	1497	4	US-09-220-132-94
27	34	5.4	3568	3	US-09-218-363-3

28	33.8	5.3	117	4	US-09-621-976-9850	Sequence 9850, Ap
29	33.8	5.3	177	4	US-09-621-976-1047	Sequence 1047, Ap
C 30	33	5.2	2417	1	US-08-011-398B-1	Sequence 1, Appli
C 31	33	5.2	2417	1	US-08-464-051-1	Sequence 1, Appli
C 32	33	5.2	2417	2	US-08-462-498-1	Sequence 1, Appli
C 33	33	5.2	2417	3	US-08-554-385-2	Sequence 2, Appli
C 34	32.8	5.2	296	4	US-09-513-999C-24392	Sequence 24392, A
C 35	32.8	5.2	1446	4	US-09-149-476-139	Sequence 139, App
36	32.6	5.1	598	3	US-09-247-155-42	Sequence 42, Appl
37	32.6	5.1	2006	4	US-09-354-123-5	Sequence 5, Appli
C 38	32.6	5.1	90541	4	US-09-759-359A-3	Sequence 3, Appli
C 39	32.6	5.1	90541	4	US-10-207-973-3	Sequence 3, Appli
40	32.4	5.1	427	4	US-09-702-705-854	Sequence 854, App
41	32.4	5.1	427	4	US-09-736-457-854	Sequence 854, App
42	32.4	5.1	427	4	US-09-614-124B-854	Sequence 854, App
43	32.4	5.1	427	4	US-09-671-325-854	Sequence 854, App
44	32.4	5.1	427	4	US-09-658-824-854	Sequence 854, App
45	32	5.0	2494	3	US-09-724-864-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-979-847B-102
; Sequence 102, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; BESEME, FREDERIC
; BEDIN, FREDERIC
; PARANHOS-BACCALA, GLAUCIA
; KOMURIAN-PRADEL, FLORENCE
; JOLIVET-REYNAUD, COLETTE
; MANDRAND, BERNARD
; GARSON, JEREMY
; TUBE, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSER: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/979,847B
APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-NO. 6582703-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 635 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-08-979-847B-102

Query Match 100.0%; Score 635; DB 4; Length 635;
Best Local Similarity 100.0%; Pred. No. 3.5e-202; Indels 0; Gaps 0;
Matches 635; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCTGATCTTTAACTCTCTGTTAAAGTTTGTCTCTCCAGAAATCAAACTGTAAACTA 60
DB 1 CCCTGATCTTTAACTCTCTGTTAAAGTTTGTCTCTCCAGAAATCAAACTGTAAACTA 60

QY 61 CAAATGTTCTTCAATGAGACACAGATGAGTCCATGATCAAGATCCACCGTGACCC 120
DB 61 CAAATGTTCTTCAATGAGACACAGATGAGTCCATGATCAAGATCCACCGTGACCC 120

QY 121 CTGACCGGCTGTAGCCCATCTCGATGTTAAATGACATTCAGAGCACCCCTCCCGAG 180
DB 121 CTGACCGGCTGTAGCCCATCTCGATGTTAAATGACATTCAGAGCACCCCTCCCGAG 180

QY 181 GAAATCTCAATGACACACCCCTACTATGCCCAATTCAGCGGAGAGCTTACGCGGT 240
DB 181 GAAATCTCAATGACACACCCCTACTATGCCCAATTCAGCGGAGAGCTTACGCGGT 240

QY 241 CATCAGCAACCTCCCAACAGACACTTGGGTTTCTGTTGAGAGGGGAGCTGAGAGC 300
DB 241 CATCAGCAACCTCCCAACAGACACTTGGGTTTCTGTTGAGAGGGGAGCTGAGAGC 300

QY 301 AGGACTAGCTGGATTCTTAGGCCCAACAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACT 360
DB 301 AGGACTAGCTGGATTCTTAGGCCCAACAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACT 360

QY 361 GCATCCACCTTAAACATGGGCTTGCACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
DB 361 GCATCCACCTTAAACATGGGCTTGCACTTAGCTCACACCCGACCAATCAGAGAGCTC 420

QY 421 ACTAAATGCTTAATAGGCAAAATAGGAGTAAAGAAATAGCCCAATCATCTATTGCTG 480
DB 421 ACTAAATGCTTAATAGGCAAAATAGGAGTAAAGAAATAGCCCAATCATCTATTGCTG 480

QY 481 AGACACAGCGGAGGACAGAGTCCGGATATAAACCAGGATTCAGCGCGCAACGG 540
DB 481 AGACACAGCGGAGGACAGAGTCCGGATATAAACCAGGATTCAGCGCGCAACGG 540

QY 541 CAACCCCTTTGGTCCCTCCCTTTGTATGGGGCTCTGTTTCACTCTATTTCACCTCT 600
DB 541 CAACCCCTTTGGTCCCTCCCTTTGTATGGGGCTCTGTTTCACTCTATTTCACCTCT 600

QY 601 ATTAATCTTGCACTGAAAAAAGAAAAA 635
DB 601 ATTAATCTTGCACTGAAAAAAGAAAAA 635

RESULT 2
US-08-979-847B-108
; Sequence 108, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; BESENE, FREDERIC
; BEDIN, FREDERIC
; PARANHOS-BACCALA, GLAUCIA
; KOMURIAN-PRADEL, FLORENCE
; JOLIVET-REYNAUD, COLETTE
; MANDRAND, BERNARD
; GARSON, JEREMY
; TUBE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
; THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA

COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-No. 6582703-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 108:
SEQUENCE CHARACTERISTICS:
LENGTH: 1329 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 108:
US-08-979-847B-108

Query Match 85.8%; Score 545; DB 4; Length 1329;
Best Local Similarity 92.4%; Pred. No. 7.3e-172; Indels 0; Gaps 0;
Matches 572; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 1 CCCTGATCTTTAACTCTCTGTTAAAGTTTGTCTCTCCAGAAATCAAACTGTAAACTA 60
DB 257 CCCTGATCTTTCAACTCTCTGTTAAAGTTTGTCTCTCCAGAAATCAAACTGTAAACTA 316

QY 61 CAAATGTTCTTCAATGAGACACAGATGAGTCCATGATCAAGATCCACCGTGACCC 120
DB 317 CAAATGTTCTTCAATGAGACACAGATGAGTCCATGATCAAGATCCACCGTGACCC 376

QY 121 CTGACCGGCTGTAGCCCATCTCGATGTTAAATGACATTCAGAGCACCCCTCCCGAG 180
DB 377 CTGACCGGCTGTAGACTATGCTCTGATGTTAAATGACATTCAGAGCACCCCTCCCGAG 436

QY 181 GAAATCTCAATGACACACCCCTACTATGCCCAATTCAGCGGAGAGCTTACGCGGT 240
DB 437 GAAATCTCAATGACACACCCCTACTACTCAATTCAGTAGAGAGAGCTTACGCGGT 496

QY 241 CATCAGCAACCTCCCAACAGCACTTGGGTTTCTGTTGAGAGGGGAGCTGAGAGAC 300
DB 497 TGTCCGCAACCTCCCAACAGCACTTGGGTTTCTGTTGAGAGGGTGGAGCTGAGAGAC 556

QY 301 AGGACTAGCTGGATTCTTAGGCCCAACAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACT 360
DB 557 AGGACTAGCTGGATTCTTAGGCCCTGACTAAGAAATCCCAAGCTTACTGGGAAGGTGACC 616

QY 361 GCATCCACCTTAAACATGGGCTTGCACCTAGCTCACACCCGACCAATCAGAGAGCTC 420
DB 617 GCATCCATCTTTAAACATGGGCTTGCACCTAGCTCACACCCGACCAATCAGAGAGCTC 676

QY 421 ACTAAATGCTTAATAGGCAAAATAGGAGTAAAGAAATAGCCCAATCATCTATTGCTG 480
DB 677 ACTAAATGCTTAATAGGCAAAATAGGAGTAAAGCAATAGCCCAATCATCTATTGCTG 736

QY 481 AGAGCACAGCGGAGGAGCAAGGATCGGATATAAACCAGGCAATTCAGCGCGGCAACGG 540
DB 737 AGAGCACAGCGGAGGAGCAAGGATTCGGATATAAATCAAGCAATTCAGCGCAACAG 796

QY 541 CAACCCCTTTGGGTCCCTCCCTTTGTATGGGGCTCTGTTTCACTCTATTTCACCTCT 600
DB 797 CAACCCCTTTGGGTCCCTCCCTTTGTATGGGGCTCTGTTTCACTCTATTTCACCTCT 856

QY 601 ATTAATCTTGCAACTGAA 619

Db 857 ATTAATCATGCACTGCA 875
|||||
RESULT 3
US-09-175-928-3
; Sequence 3, Application US/09175928A
; Patent No. 6312921
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Mi, Sha
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 6006B.A172A
; CURRENT APPLICATION NUMBER: US/09/175,928A
; CURRENT FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2946
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-175-928-3

Query Match 84.1%; Score 533.8; DB 3; Length 2946;
Best Local Similarity 91.5%; Pred. No. 6.8e-168;
Matches 581; Conservative 0; Mismatches 42; Indels 12; Gaps 1;
QY 1 CCTGTATCTTTAACTCTCTTGTGTTAGTTGTCTTCCAGAAATCAAACTGTAAACTA 60
Db 2323 CCTGTATCTTTAACTCTCTTGTGTTAGTTGTCTTCCAGAAATCGAGCTGTAAACTA 2382
QY 61 CAAATTTGTTCTTCAAATGGAGCACAGATGGAGTCCATGACTCAAGATCCACCGTGGACCC 120
Db 2383 -----CAAATGGAGCCCAAGATGTCAGTCCCAAGACTAAGATCTAGCGCAGACCC 2430
QY 121 CTGGACGGGCTGCTAGCCCATGCTCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
Db 2431 CTGGACGGGCTGCTAGCCCATGCTCGATGTTAATGACATCAAGGCACCCCTCTCTGAG 2490
QY 181 GAAATCTCACTGCACAACTCTACTATGCCCCAATTCAGCGGGAAGCAGTTAGACGGT 240
Db 2491 GAAATCTCACTGCACAACTCTACTAGCCCCAATTCAGCGGGAAGCAGTTAGACGGT 2550
QY 241 CATCAGCCCACTCCCCAACAGACACTTGGGTTTTCTCTGTTGAGAGGGGGGACTGAGAGAC 300
Db 2551 CGTCGGCCCACTCCCCAACAGACACTTGGGTTTTCTCTGTTGAGATGGGGGACTGAGAGAC 2610
QY 301 AGACTAGCTGGAATTCCTTAGGCCAACAGAAATCCCTAAGCTAGCTAGCTGGGAAGGTGACT 360
Db 2611 AGACTAGCTGGAATTCCTTAGGGCTGACTAAGAAATCCCTAAGCTAGCTAGCTGGGAAGGTGACC 2670
QY 361 GCATCCACTCTAAACATGGGCTTGCACCTAGCTACACCCGACCAATCAGAGAGCTC 420
Db 2671 ACATCCACTCTTAAACACGGGGCTTGCACCTAGCTACACCTGACCAATCAGAGAGCTC 2730
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTCCCTG 480
Db 2731 ACTAAATGCTAATTAGGCAAAACAGGAGGTAAAGAAATAGCAATCATCTATTCCCTG 2790
QY 481 AGAGCAGCGGGAGGACAAAGATCGGGATATAAACCAGGCAATTCGAGCGGCAACGG 540
Db 2791 AGAGCAGCAGGAGGACAAATGATCGGGATATAAACCAGGCAATTCGAGCGGCAACGG 2850
QY 541 CAACCCCTTTGGGTCCCTCCCTTTGTATGGGGCTCTGTTTCACTCTATTTCACCT 600
Db 2851 CAACCCCTTTGGGTCCCTCCCTTTGTATGGGGAGCTCTGTTTTCATGCTATTTCACCT 2910

QY 601 ATTAATCTTGCACCTGAAAAAAGAAAAAAGAAAAA 635
Db 2911 ATTAATCTTGCACCTGCAAAAAAAGAAAAAAGAAAAA 2945
RESULT 4
US-08-686-878A-50
; Sequence 50, Application US/08686878A
; Patent No. 5708157
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: LaVallie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Evans, Cheryl
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/686,878A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-686-878A-50

Query Match 36.8%; Score 233.6; DB 1; Length 279;
Best Local Similarity 88.1%; Pred. No. 4.1e-68;
Matches 245; Conservative 1; Mismatches 32; Indels 0; Gaps 0;
QY 358 ACTGCATCCACCTTAAACATGGGGCTTGCACTTAGCTCACACCCGACCAATCAGAGAG 417
Db 1 RCACATCCACCTTTAAACACGGGGNTTGCANANAGATNACACTTGACCAATCAGAGAG 60
QY 418 CTCACCTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGC 477
Db 61 NTCANTAAATGATNATTTGGGCAAAAGAGGTAAGAAATAGCAATCATCTATTGC 120
QY 478 CTGAGAGCAGCGGGAGGACAGGATCGGGATATAAACCAGGCAATTCGAGCCCGCAA 537
Db 121 CTGAGAGCAGCAGGAGGACAAATGATCGGGATATAAACCAGGCAATTCGAGCCCGCAA 180
QY 538 CGGCAACCCCTTTGGGTCCCTCCCTTTGTATGGCGCTCTGTTTCACTCTATTTCAC 597
Db 181 CGGCAACCCCTTTGGGTCCCTCCCTTTGTATGGGAGCTGTTTTCATGCTATTTCAN 240

QY 598 TCTATTAAATCTTGCACTGCAAAAAAAAAAAAAAAAAAAAA 635
Db 241 TTTATTAAATNTTGCACCTGCAAAAAAAAAAAAAAAAAAAAA 278

RESULT 5
US-08-721-489-4
; Sequence 4, Application US/08721489
; Patent No. 5786465
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: LaVallie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/721,489
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-721-489-4

Query Match 36.8%; Score 233.6; DB 1; Length 279;
Best Local Similarity 88.1%; Pred. No. 4.1e-68;
Matches 245; Conservative 1; Mismatches 32; Indels 0; Gaps 0;
QY 358 ACTGCATCCACCTTAAACATGGGGCTTGCAACTTACTCACCACCGCAACATCAGAGAG 417
Db 1 RCCACATCCACCTTAAACACGGGNTTGAANAAGATNACACTTGACCAATCAGAGAG 60
QY 418 CTCACATAATGCTAATTAGCAAAATAGAGGTAAGAAATAGCCCAATCATCTATTGC 477
Db 61 NTCANTAAATGATNATNGGCAAAACAGAGGTAAGAAATAGCCCAATCATCTATTGC 120
QY 478 CTGAGACACAGCGGGAGGACAGGATCGGATATAAACCAGGCATTGAGCGGCA 537
Db 121 CTGAGACACAGAGGAGGACAAATGATCGGATATAAACCAGGTTTNGAGCGGCA 180
QY 538 CGGCAACCCCTTTGGGTCCTCCCTTTGATGGGCTCTGTCTTCTACTCTATTTCAC 597
Db 181 CGGCAACCCCTTTGGGTCCTCCCTTTGATGGGAGCTNTGTTTCATGCTATTTCAN 240
QY 598 TCTATTAAATCTTGCACTGCAAAAAAAAAAAAAAAAAAAAA 635

Db 241 TTTATTAAATNTTGCACCTGCAAAAAAAAAAAAAAAAAAAAA 278

RESULT 6
US-09-621-976-9366/c
; Sequence 9366, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 9366
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-9366

Query Match 31.0%; Score 196.6; DB 4; Length 456;
Best Local Similarity 84.9%; Pred. No. 1.3e-55;
Matches 264; Conservative 0; Mismatches 24; Indels 23; Gaps 3;
QY 292 CTGAGAGACAGGACTAGCTGGATTTCTTAGGCCAACGAAAGATCCCTAAGCCCTAGCTGGG 351
Db 307 CTGAGAGACAGGACTAGCTGGATTTCTTAGGCCAACGAAAGATCCCTAAGCCCTAGCTGGG 248
QY 352 AAGGTGACTGCAATCCACCTCTAAACATGGGGCTTGCACTTAGCTCACCACCGACCAATC 411
Db 247 AAGGTGACCACATCCACCTTTAAGCAGGGGGCTTGCACTTAGCTCACCACCGACGATC 188
QY 412 -----AGAGAGCTCACTAAATGCTAATTAGCAAAATAGGAGGTAAAGAAATAG 462
Db 187 AGGTAGTAAAGCGAGCTCACTAAATGCTAATTAGGCTAAACAGAGGATTAAC-CAATAG 129
QY 463 CCAATCATCTATTGCTCTGAGAGACACGCGGGAGGACAAGGATCGGATATAAACCCAGG 522
Db 128 CCATCATCTATCTCTCTGAGACACAGCAGGAGGAGCAATGATAGGATATAAAC----- 74
QY 523 CATTCGAGCGGCAACGGCAACCCCTTTGGGTCCCTCCCTTTGATGGGCGCTCTGTT 582
Db 73 -----CTGGCAACGGCTCTCTGGGTCCCTCTCTGGGTCCCTTTGTATGGGAGCTCTGTT 22
QY 583 TTCACCTCTATT 593
Db 21 TTCACCTCTATT 11

RESULT 7
US-08-691-563C-46
; Sequence 46, Application US/08691563C
; Patent No. 6001987
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glauca PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Collette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia

```

; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/691,563C
; APPLICATION NUMBER: US/08/691,563C
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38598
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1859 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-691-563C-46

Query Match 20.2%; Score 128; DB 3; Length 1859;
Best Local Similarity 62.7%; Pred.No.3e-32;
Matches 237; Conservative 0; Mismatches 130; Indels 11; Gaps 2;

QY 1 CCCTGTATCTTTAACTCTTTAGTTTGTCTCTCCAGAATCAAACTGTAAGACTA 60
Db 1446 CCCTGTATTTTAACTCTTTGTCAATTGTTTCTCTAGGATCGAGGCCATCAGCTA 1505

QY 61 CAAATTGTTCTTCAAATGAGACACCATGGAGTCCATGACTAAGATCCACCGTGAACCC 120
Db 1506 CAGATGGTCTTACAAATGGAACCCCAATGAGCTCAACTATCAACTTCTACTGAGACCC 1565

QY 121 CTGACCGGGCTCTAGCCCCATGCTCGGATGTTAATGACATTCGAAGCACCCCTCCCGAG 180
Db 1566 CTAGACCAACCCCTGGCCC-----TTTCACTGGCCTAAAGATTCCCTCTGGAG 1616

QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
Db 1617 GACACTACCACCTGCAGGGGCCCATCTTTGGCCCTATCCAGAAGGAAGTAGCTAGAGCAGT 1676

QY 241 CATTAGCCAACTCCCAACAGACACTTGGGTTTTCCTGTTGAGAGGGGGAAGTGAAGAC 300
Db 1677 CATTTGCCAA--TTCCCAAGAGCAGCTGGGGTGTCCCGTTTTAGAGTGGGGAATTGAGAGGT 1734

QY 301 AGGACTAGCTGGATTTCTTAGGCCAACGAAGAATCCCTTAAGCCCTAGCTGGGAAGGTGACT 360
Db 1735 GAAGCCAGCTGGACTTCTGGGTGGGTGGGACTTTGTGTCTAGCTAAG 1794

QY 361 GCATCCACCTCTAAACAT 378
Db 1795 GATTGTAAATGCAACAT 1812

```

```

: TITLE OF INVENTION: THERAPEUTIC PURPOSES
: NUMBER OF SEQUENCES: 92
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Oliff & Berridge
: STREET: 700 South Washington Street, Suite 300
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22314
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/374,766
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/691,563
: FILING DATE: 02-AUG-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Berridge, William P.
: REGISTRATION NUMBER: 30,024
: REFERENCE/DOCKET NUMBER: WPB 39588
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-836-6400
: TELEFAX: 703-836-2787
: INFORMATION FOR SEQ ID NO: 46:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1859 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cdna
: US-09-374-766-46

Query Match 20.2%; Score 128; DB 4; L
Best Local Similarity 62.7%; Pred. No. 3e-32;
Matches 237; Conservative 0; Mismatches 130;

Qy 1 CCTGTGATCTTTAACTCCTCTGTTAAGTTTGTCTCTTCCAA
Db 1446 CCTGTATTTTAACTCCTCTGTCAAAATTTGTTCTCTTA
Qy 61 CAAATGTGTTCTTCAAATGAGCACCAGATGGAGTCCATGA
Db 1506 CAGATGCTTACAAATGAA CCCCAAATGAGCTCAACTA
Qy 121 CTGGACGGGCTGTAGCCCATGCTCCGATGTTAATGACA
Db 1566 CTAGACCAACCCCTGGCCC-----TTTCACTGGCC
Qy 181 GAAATCTCACTGCACACCCCTACTATGCCCAATTGAG
Db 1617 GACACTACCACTGCAGGGGCCCATCTTTGCCCTATCCAG
Qy 241 CATCAGCAACCTCCCCCAACAGACACTTGGGTTTTCCTGTT
Db 1677 CATTGCCCAA--TTCCACAGAGAGCTGGGGTGTCCCGTT
Qy 301 AGGACTAGCTGGATTTCTTAGGCCAACGAAGAATCCCTAA
Db 1735 GAAGCCAGCTGGACTTCTGGGTGGGTGGGACTTGGAGAG
Qy 361 GCATCCACCTCTAAACAT 378
Db 1795 GATTGTAATGCAACAAT 1812

RESULT 9
US-08-979-847B-42
: Sequence 42, Application US/08979847B
: Patent No. 6587303

```

ATTORNEY: BELMONT, MANDARINO
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC

GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
BESME, FREDERIC
BEDIN, FREDERIC
PARANHOS-BACCALA, GLAUCIA
KOMURIAN-PRADEL, FLORENCE
JOLIVET-REYNAUD, COLETTE
MANDRAND, BERNARD
GABSON, JEREMY
TUBE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979.847B
FILING DATE: 26-No. 6582703-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30, 024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1859 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-08-979-847B-42

Query Match 20.2%; Score 128; DB 4; Length 1859;
Best Local Similarity 62.7%; Pred. No. 3e-32;
Matches 237; Conservative 0; Mismatches 130; Indels 11; Gaps 2;
Db 1446 CCCTGTATCTTAACTCCTTGTAACTTTGCTCTCTCCAGAACTCAAACTGTAATACTA 60
1506 CAGATGTCTTACAAATGGAAACCCCAATGAGTCAACTATCACTTCTACTGAGGACCC 1565
121 CTGACCGCGCTGTAGCCCATGCTCCGATGTTAATGACATTTGAAGCACCCTCCCGAG 180
1566 CTAGACCAACCCCTGGGCC-----TTTCACTGGCCCTAAAGAGTTCCTCTCTGGAG 1616
181 GAATCTCAATGTCACAAACCCCTACTATGCCCAATTCAGCGGAAGCAGTTAGAGCGGT 240
1617 GACACTACCTAGCGGGCCCCATCTTTGGCCCCATCCAGNAGNAGTAGTAGAGGT 1676
241 CATCAGCAACCTCCCAACAGCACTGGGTTTTCTGTTGAGAGGGGGGACTGAGAGAC 300
1677 CATTTGCCAA--TTCCCAAGAGCAGCTGGGGTGTCCCGTTTAGAGTGGGATTGAGAGGT 1734
301 AGGACTAGTGGATTCTCTAGGCCAACAGAAATCCCTAAGCCTAGCTGGGAAGTCACT 360

Db 1735 GAAGCAGCTGGACTTCTGGTGGTGGGACTTGGAGAACTTTTGTCTAGCTAAAG 1794
QY 361 GCATCCACCTCTAAACAT 378
Db 1795 GATTGTAATGCAACAAT 1812
RESULT 10
US-09-078-294-4
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 80246
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match 19.8%; Score 125.8; DB 3; Length 80246;
Best Local Similarity 67.2%; Pred. No. 1.8e-30;
Matches 213; Conservative 0; Mismatches 92; Indels 12; Gaps 2;
QY 1 CCCTGTATCTTAACTCCTTGTAACTTTGCTCTCTCCAGAACTCAAACTGTAATACTA 60
Db 5888 CCCTGTATTTTAACTCCTTGTGTCATAATTTGTTCTCTAGGATCGAGGCCATCAAGCTA 58947
QY 61 CAAATTGTTCTCAATGGAGCAGACGAGTCCATGACTAAGATCAACCGTGGACCC 120
Db 58948 CAGATGATCTTACAAATGTAACCCCAATGAGTCAACTTACAACTTCTCTGAGGACCC 59007
QY 121 CTGACCGCGCTCTAGCCCATGCTCCGATGTTAATGACATTTGAAGGCACCCCTCCCGAG 180
Db 59008 CTGACCGGACCCCTGGGCC-----TTTCAATGGCCCTAAAGAGTCCCTCTGGAG 59058
QY 181 GAAATCTCACTGCACAAACCCCTACTATGCCCCCAATTCAGCGGAAGCAGTTAGAGCGGT 240
Db 59059 GACACTACCACTGCAGGGGCCCTTCTTACCCCTATCCAGCAGGAAAGTAGCTACAGCGGT 59118
QY 241 CATCAGCAACCTCCCAACAGCACTGGGTTTTCTGTTGAGAGGGGGGACTGAGAGAC 300
Db 59119 CATCGCCAAA---TCCCAACAGCAGCTGGGGTGTCTCTGTTGAGAGGGGGATTGAGAGGT 59175
QY 301 AGGACTAGCTGGATTTC 317
Db 59176 GAAGCCAGCTGGGCTTC 59192

RESULT 11
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA

```
; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

Query Match          19.8%; Score 125.8; DB 3; Length 80595;
Best Local Similarity 67.2%; Pred. No. 1.8e-30;
Matches 213; Conservative 0; Mismatches 92; Indels 12; Gaps 2;

QY 1 CCTGTATCTTTAACTCCTTGTGTTAACTTGTCTCTTCCAGAAATCAAACTGTAAACTA 60
Db 59121 CCTGTATCTTTAACTCCTTGTGTTAACTTGTCTCTTCCAGAAATCAAACTGTAAACTA 59180

QY 61 CAAATGTTCTTCAAAATGGAGCACCAGATGGAGTCCATGACTAAGATFCCACCGTGGACCC 120
Db 59181 CAGATGATCTTACAAATGTTAACTCCCAATGAGTCACTAACAATCTTCTGCTGAGGACCC 59240

QY 121 CTGGACCGGCTGTAGCCCATGCTCCGATGTTAATGACATTTGAAGGCAACCCCTCCCGAG 180
Db 59241 CTGGACCGGCTGTAGCCCATGCTCCGATGTTAATGACATTTGAAGGCAACCCCTCCCGAG 59291

QY 181 GAAATCTCACTGCACAAACCCCTACTATGCCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
Db 59292 GACACTACCACTGAGGGCCCCCTTCTTCAACCCCTATCCAGCAGGAAGTAGCTACAGCGGT 59351

QY 241 CATCAGCAACCTCCCCCAACAGACACTTGGGTTTTTCTGTTTGAAGGGGGGAGCTGAGAGAC 300
Db 59352 CATCGCCAAA---TCCCAACAGAGCTGGGGTGTCTGTTTGAAGGGGGGAGTTGAGAGGT 59408

QY 301 AGGACTAGCTGGATTC 317
Db 59409 GAAGCCAGCTGGGCTTC 59425

RESULT 12
US-08-691-563C-58
; Sequence 58, Application US/08691563C
; Patent No. 6001987
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glaucia PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,563C
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38588
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1722 base pairs
```

```
; TYPE: nucleotide
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-691-563C-58

Query Match          19.7%; Score 125; DB 3; Length 1722;
Best Local Similarity 69.2%; Pred. No. 2.8e-31;
Matches 204; Conservative 0; Mismatches 80; Indels 11; Gaps 2;

QY 1 CCTGTATCTTTAACTCCTTGTGTTAACTTGTCTCTTCCAGAAATCAAACTGTAAACTA 60
Db 796 CCTGTATCTTTAACTCCTTGTGTTAACTTGTCTCTTCCAGAAATCAAACTGTAAACTA 855

QY 61 CAAATGTTCTTCAAAATGGAGCACCAGATGGAGTCCATGACTAAGATFCCACCGTGGACCC 120
Db 856 CAGATGTTCTTACAAATGGAAACCCCAATGAGTTCAACTAACAACTTCTTCAAAAGGACCC 915

QY 121 CTGGACCGGCTGTAGCCCATGCTCCGATGTTAATGACATTTGAAGGCACCCCTCCCGAG 180
Db 916 CTGGAAAGCATCCACTGGCACTTCC-----ACTAGCCTAGAGATTCCCTCTCTGAA 965

QY 181 GAAATCTCACTGCACAAACCCCTACTATGCCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
Db 966 GACACTACCACTGAGGGCCCCCTTCTTGGCCCTATCCAGCAGGAAGTAGCTAGAGCGGT 1025

QY 241 CATCAGCAACCTCCCCCAACAGACACTTGGGTTTTTCTGTTTGAAGGGGGGAGCTGA 295
Db 1026 CATCGGCCAA-AATCCCAACAGAGTTGGGGTGTCTGTTTGAAGGGGGGAGTTGA 1079

RESULT 13
US-09-374-766-58
; Sequence 58, Application US/09374766
; Patent No. 6579526
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glaucia PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/374,766
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,563
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38588
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 58:
```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1722 base pairs
; TYPE: nucleotide
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-374-766-58

Query Match          19.7%; Score 125; DB 4; Length 1722;
Best Local Similarity 59.2%; Pred. No. 2.8e-31;
Matches 204; Conservative 0; Mismatches 80; Indels 11; Gaps 2;

QY 1 CCTGTATCTTTAACTCCTCTGTTAAAGTTTGTCTCTTCCAGATCAAAACTGTAAACTA 60
Db 796 CCTGTATCTTTAAAGTTTGTCTCTTCCAGATCAAAACTGTAAACTA 855
QY 61 CAAATGTTCTTCAATGAGACCAAGATGAGTCCATGACTAAGATCCACCGTGGACCC 120
Db 856 CAGATGGTCTTACAAATGGAACCCCAATGAGTTCAACTAACACTTCTACCAAGGACCC 915
QY 121 CTGGACCGGCTGCTAGCCCATGCTCGATGTTAAATGACATTGAAGGCACCCCTCCCGAG 180
Db 916 CTGGAAGCATCCACTGGCACTTC-----ACTAGCTAGAGATCCCTCTGGAA 965
QY 181 GAAATCTCAACTCACAAACCCCTACTATGCCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
Db 966 GACACTAGAACTGACGGGCCCTTCTTTGCCCTATCCAGCAGGAAGTAGTAGAGCGGT 1025
QY 241 CATCAGCAACCTCCCAACAGCACTGGGTTTCTCTGTTGAGAGGGGGAAGTGA 295
Db 1026 CATCGGCCAA-ATTCCCAACAGCAGTTGGGGTGTCTCTGTTAGAGGGGGGATTGA 1079

RESULT 14
US-08-979-847B-54
; Sequence 54, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; BESEME, FREDERIC
; BEDIN, FREDERIC
; PARANHOS-BACCALA, GLAUCIA
; KOMURIAN-PRADEL, FLORENCE
; JOLIVET-REYNAUD, COLETTE
; MANDRAND, BERNARD
; GARSON, JEREMY
; TUKE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
; THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979 847B
; FILING DATE: 26-Nov-95
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BERRIDGE, WILLIAM P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 35046A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400

```

```

; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1722 base pairs
; TYPE: nucleotide
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-979-847B-54

Query Match          19.7%; Score 125; DB 4; Length 1722;
Best Local Similarity 59.2%; Pred. No. 2.8e-31;
Matches 204; Conservative 0; Mismatches 80; Indels 11; Gaps 2;

QY 1 CCTGTATCTTTAACTCCTCTGTTAAAGTTTGTCTCTTCCAGATCAAAACTGTAAACTA 60
Db 796 CCTGTATCTTTAAAGTTTGTCTCTTCCAGATCAAAACTGTAAACTA 855
QY 61 CAAATGTTCTTCAATGAGACCAAGATGAGTCCATGACTAAGATCCACCGTGGACCC 120
Db 856 CAGATGGTCTTACAAATGGAACCCCAATGAGTTCAACTAACACTTCTACCAAGGACCC 915
QY 121 CTGGACCGGCTGCTAGCCCATGCTCGATGTTAAATGACATTGAAGGCACCCCTCCCGAG 180
Db 916 CTGGAAGCATCCACTGGCACTTC-----ACTAGCTAGAGATCCCTCTGGAA 965
QY 181 GAAATCTCAACTCACAAACCCCTACTATGCCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
Db 966 GACACTAGAACTGACGGGCCCTTCTTTGCCCTATCCAGCAGGAAGTAGTAGAGCGGT 1025
QY 241 CATCAGCAACCTCCCAACAGCACTGGGTTTCTCTGTTGAGAGGGGGAAGTGA 295
Db 1026 CATCGGCCAA-ATTCCCAACAGCAGTTGGGGTGTCTCTGTTAGAGGGGGGATTGA 1079

RESULT 15
US-09-513-999C-24228
; Sequence 24228, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 24228
; LENGTH: 149
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 127
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 129
; OTHER INFORMATION: n=a, g, c or t
; NAME/KEY: misc_feature
; LOCATION: 135
; OTHER INFORMATION: m=a or c
; OTHER INFORMATION: m=a or c
US-09-513-999C-24228

Query Match          17.4%; Score 110.6; DB 4; Length 149;
Best Local Similarity 90.1%; Pred. No. 3.9e-27;

```

```
Matches 127; Conservative 2; Mismatches 11; Indels 1; Gaps 1;
QY 1 CCTGTATCTTTAACTCCCTTGTAGTTTGTCTCTCCAGAAATCAAAATGTAAACTA 60
Db 5 CTCTGTATCTTTAACTCCCTTGTAGTTTGTCTCTCCAGAAATCAAAATGTAAACTA 64
QY 61 CAAATGTTCTTCAAAATGGAGCACCAGATGAGTCCATGACTAAGATCCACCGTGGACCC 120
Db 65 CAAATGTTCTTCAAAATGGAGCACCAGATGAGTCCATGACTAAGATCTTACTGGGACCT 124
QY 121 CTGG-ACCGGCTGCTAGCCC 140
Db 125 CTRGNACCGGCTGCTAGCCC 145
RESULT 16
US-09-120-653D-1
; Sequence 1, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
; APPLICANT: YOON, JI-WON
; APPLICANT: JUN, HEE-SOOK
; APPLICANT: PARK, HAE-JOON
; APPLICANT: AHN, JONG-SEONG
; APPLICANT: HA, YOUNG-JU
; APPLICANT: CHUNG, SOO-IL
; TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
; FILE REFERENCE: 98-338
; CURRENT APPLICATION NUMBER: US/09/120,653D
; CURRENT FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: KR 98-10108
; PRIOR FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 1
; LENGTH: 3910
; TYPE: DNA
; ORGANISM: DIABETES-SPECIFIC ENDOGENOUS RETROVIRUS ERV-9
US-09-120-653D-1
Query Match 12.0%; Score 76.4; DB 3; Length 3910;
Best Local Similarity 61.4%; Pred. No. 8.8e-15;
Matches 178; Conservative 0; Mismatches 101; Indels 11; Gaps 3;
QY 1 CCTGTATCTTTAACTCCCTTGTAGTTTGTCTCTCCAGAAATCAAAATGTAAACTA 60
Db 3194 CCCAGCATTTTAAACGTTCTTGTCAAATTTGT-TTGTCTAGATTCGAGGCCTTCAAGCTA 3252
QY 61 CAAATGTTCTTCAAAATGGAGCACCAGATGAGTCCATGACTAAGATCCACCGTGGACCC 120
Db 3253 CAGATGGTCTTACATATCAAAACCCCGAATAATTTCAACTAACTTCTTACGGAGGACAC 3312
QY 121 CTGGACCGGCTGCTAGCCCTGCTCGATGTTAAATGACATTCGAAGCACCCTCCGAG 180
Db 3313 CTGGACTAACGCTGGCAGTTACCC-----TGGCCTAGAGAGTTCCCTCTGGAAG 3363
QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCCAAATTCAGCGGAAGCAGTTAGAGCGGT 240
Db 3364 GTCACTACAACTGCAAGGCCCTTTTCGGCCCTATCCATCAGGAGCTACTAGAACAGT 3423
QY 241 CATCAGCCAACTCCCAACAGACACTTGGTTTTCTGTGTTGAGGGGGG 290
Db 3424 CCTCGGCC-ATAATGCAACAGAGTTGGAGTGTCTGTTGATTGAGGGG 3472
RESULT 17
US-09-621-976-11057
; Sequence 11057, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
```

```
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 11057
; LENGTH: 492
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-11057
Query Match 11.6%; Score 73.4; DB 4; Length 492;
Best Local Similarity 60.3%; Pred. No. 2.3e-14;
Matches 164; Conservative 0; Mismatches 96; Indels 13; Gaps 2;
QY 24 TAAGTTTGTCTCTTCCAGAAATCAAAATGTAAACTACAAATTTGTTCTTCAAAATGGAGCA 83
Db 95 TAATCAAGTCTAATCAAGGATCAAGCCATCAAGCTGCAATGGACTTACAAATGGAAACC 154
QY 84 CCAGATGGAGTCCATGACTAAGATCCACCGTGGACCCCTCGACCGGCTGTAGCCCATG 143
Db 155 TCAATGAGCTCAACTCAGGGCTTCTACTGAGAACCCCTGGATCCACCTGTGTGCTCCTC 214
QY 144 CTCGGATGTTAATGACATTGAAGGCACCCCTCCCGAGGAATCTCAACTGCACAAACCCCT 203
Db 215 -----ACTAGCCTAAAGTTCCCTCTCGAGGACACCACTGCAGGGCCCT 264
QY 204 ACTATGCCCAATTCAGCGGAGCAGTTAGAGCGGTCAATCAGCAACCTCCCAACAGC 263
Db 265 TCTTCAACCTTACCATCAGGAAGTAGCCAGAACG---ACTGTGCCAGTTCCCAACAGC 321
QY 264 ACTTGGGTTTTCTGTGAGAGGGGGGACTGAG 296
Db 322 AATTGGTGTCTCTGTAGAGGGGGGACTGAG 354
RESULT 18
US-09-513-999C-21394
; Sequence 21394, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59 US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 21394
; LENGTH: 443
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 102
; OTHER INFORMATION: k=g or t
US-09-513-999C-21394
Query Match 11.3%; Score 71.8; DB 4; Length 443;
Best Local Similarity 89.2%; Pred. No. 7.5e-14;
Matches 99; Conservative 1; Mismatches 8; Indels 3; Gaps 2;
QY 510 ATATAAACCCAGGCAATTCGAGCGCGCAACGGCA-ACCCCTTTGGGTCCCTCCCTTTGT 568
Db 3 ATATAAACCCAGGCAATTCGAGCTGGCAACAGAGCCCCCTTTGGGTCCCTCCCTTTGT 62
QY 569 ATGGGCGCTCTGTGTTTTTCACTCTATTTCACCTCTATTAAATCTTCAAGTAA 619
```

Db 63 ATGGGAG--CTGTTTCATGCTATTTCACCTCTATTAAATCTGCAACTGCA 111

RESULT 19

US-08-979-847B-105

; Sequence 105, Application US/08979847B

; Patent No. 6582703

; GENERAL INFORMATION:

APPLICANT: PERRON, HERVE

BESEME, FREDERIC

BEDIN, FREDERIC

PARANHOS-BACCALA, GLAUCIA

KOMURIAN-PRADEL, FLORENCE

JOLIVET-REYNAUD, COLETTE

MANDRAND, BERNARD

GARSON, JEREMY

TUKE, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL

THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210

CORRESPONDENCE ADDRESS:

ADDRESSEE: OLIFF & BERRIDGE, PLC

STREET: P.O. BOX 19928

CITY: ALEXANDRIA

STATE: VA

COUNTRY: USA

ZIP: 22320

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/979,847B

FILING DATE: 26-No. 6582703-1997

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BERRIDGE, WILLIAM P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: WPB 39046A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 105:

SEQUENCE CHARACTERISTICS:

LENGTH: 1481 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 105:

US-08-979-847B-105

Query Match 10.8%; Score 68.4; DB 4; Length 1481;

Best Local Similarity 87.2%; Pred. No. 2.2e-12;

Matches 75; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 CCCTGTATCTTTAACTTCTTGTAAAGTTTGTCTTCCAGAAATCAAACTGTAAACTA 60

Db 1396 CCCTGTATCTTTAACTTCTTGTAAAGTTTGTCTTCCAGAAATTTGAAGCTGTAAAGCTA 1455

Qy 61 CAATATGTTCTTCAAAATGGAGCACA 86

Db 1456 CAGATGGTCTTTACAAATGGAAACCCA 1481

RESULT 20

US-09-513-999C-13878

; Sequence 13878, Application US/09513999C

; Patent No. 6783961

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Duclert, A.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

; Patent No. 6783961

; FILE REFERENCE: 59.US2.REG

; CURRENT APPLICATION NUMBER: US/09/513,999C

; CURRENT FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/122,487

; PRIOR FILING DATE: 1999-02-26

; NUMBER OF SEQ ID NOS: 36681

; SOFTWARE: Patent.pm

; SEQ ID NO 13878

; LENGTH: 342

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

NAME/KEY: misc_feature

LOCATION: 88

OTHER INFORMATION: y=c or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: 89

OTHER INFORMATION: k=g or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: 92

OTHER INFORMATION: y=c or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: 103

OTHER INFORMATION: k=g or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: 109

OTHER INFORMATION: s=g or c

FEATURE:

NAME/KEY: misc_feature

LOCATION: 277

OTHER INFORMATION: y=c or t

US-09-513-999C-13878

Query Match 6.5%; Score 41.4; DB 4; Length 342;

Best Local Similarity 62.4%; Pred. No. 0.00095;

Matches 63; Conservative 1; Mismatches 37; Indels 0; Gaps 0;

Qy 40 AGAATCAAAACTGTAAAACTTACAAATTTCTTCAAAATGGAGCACCAGATGGAGTCCATG 99

Db 190 AGGATCAAGCCATCAAGCTACAAATGATCTTACAAATGGAACTCAATGAGCTCAGCT 249

Qy 100 ACTAAGATCCACCGTGGACCCCTGGACCCGCGCTGTAGCCC 140

Db 250 CACGGCTTCTACCGAGGACCCCTGGATYAACCGCGTGTCC 290

RESULT 21

US-10-140-002-541

; Sequence 541, Application US/10140002

; Patent No. 6725730

; GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Beresini, Maureen

APPLICANT: DeForge, Laura

APPLICANT: Desnoyers, Luc

APPLICANT: Filvaroff, Ellen

APPLICANT: Gao, Wei-Qiang

APPLICANT: Garritsen, Mary E.

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Sherwood, Steven

APPLICANT: Smith, Victoria

APPLICANT: Stewart, Timothy A.

APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C59
CURRENT APPLICATION NUMBER: US/10/140,002
CURRENT FILING DATE: 2002-05-06
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 541
LENGTH: 3824
TYPE: DNA
ORGANISM: Homo Sapien
US-10-140-002-541

Query Match 6.5%; Score 41.2; DB 4; Length 3824;
Best Local Similarity 61.8%; Pred. No. 0.0052;
Matches 84; Conservative 0; Mismatches 48; Indels 4; Gaps 1;

QY 489 GCGGGAGGGAACAGATCGGGATATAAACCCAGCATTCGAGCGGCGCAACGGCAACCCCC 548
DB 3693 GTGGGTGGGCGAGACAAGAGATAAAAGCAGGCTGCTGAGCCAGCAGTGACAACCCCC 3752
QY 549 TTTGGGTCCCTCCCTTTGATGGCGCTCTGTTTTCACCTCTATTTCACCTCTATTAAATC 608
DB 3753 CTCGGGTCCCTCCCGCTGGAAGCTTTGTTCTTCGCTCTTTGC-----AATTAATC 3808
QY 609 TTGCAACTGAAAAA 624
DB 3809 TTGCTACTGCCAAAA 3824

RESULT 22
US-09-120-653D-6
Sequence 6, Application US/09120653D
Patent No. 6365727
GENERAL INFORMATION:
APPLICANT: YOON, JI-WON
APPLICANT: JUN, HEE-SOOK
APPLICANT: PARK, HAE-JOON
APPLICANT: AHN, JONG-SEONG
APPLICANT: HA, YOUNG-JU
APPLICANT: CHUNG, SOO-IL
TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
FILE REFERENCE: 98-338
CURRENT APPLICATION NUMBER: US/09/120,653D
CURRENT FILING DATE: 1998-07-22
PRIOR APPLICATION NUMBER: KR 98-10108
PRIOR FILING DATE: 1998-03-24
NUMBER OF SEQ ID NOS: 33
SOFTWARE: KopatentIn 1.71
SEQ ID NO 6
LENGTH: 464
TYPE: DNA
ORGANISM: DIABETES-SPECIFIC RETROVIRUS ENDOGENOUS ERV-9
US-09-120-653D-6

Query Match 6.4%; Score 40.4; DB 3; Length 464;
Best Local Similarity 62.7%; Pred. No. 0.0025;
Matches 79; Conservative 0; Mismatches 46; Indels 1; Gaps 1;

QY 1 CCTGTATCTTTAACTCCCTTTGTTAGTTTGTCTTCCAGATCAAACTGTAATACTA 60
DB 337 CCAGCATTTTAACTGTTCTTGTCAAAATTTGT-TTGTCTAGATTCGAGGCTTCAAGCTA 395
QY 61 CAAATTTGTTCTTCAATGGAGCACCAGATGCGATGCTAAGATCCACCGTGGACCC 120
DB 396 CAGATGCTCTTACATATCAAAACCCCGAATAATTTCAACTAACTTCTTACGGAGGAC 455
QY 121 CTGGAC 126
|||||

Db 456 CTGGAC 461

RESULT 23
US-09-422-978-3579/c
Sequence 3579, Application US/094222978
Patent No. 6537751
GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Blumenfeld, Marta
APPLICANT: Chumakov, Ilya
TITLE OF INVENTION: Biallelic markers for use in constructing a high density...
FILE REFERENCE: GENSET 020CP1
CURRENT APPLICATION NUMBER: US/09/422,978
CURRENT FILING DATE: 1999-10-20
EARLIER APPLICATION NUMBER: US 09/298,850
EARLIER FILING DATE: 1999-04-21
EARLIER APPLICATION NUMBER: US 60/109,732
EARLIER FILING DATE: 1998-11-23
EARLIER APPLICATION NUMBER: US 60/082,614
EARLIER FILING DATE: 1998-04-21
NUMBER OF SEQ ID NOS: 11796
SEQ ID NO 3579
LENGTH: 47
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 24
OTHER INFORMATION: 99-6549-275 : polymorphic base A or G
US-09-422-978-3579

Query Match 6.1%; Score 38.6; DB 4; Length 47;
Best Local Similarity 87.2%; Pred. No. 0.0023;
Matches 41; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 355 GTGACTGCATCCACCTCTAAACATGGGGCTTGCACCTAGCTACAC 401
DB 47 GTGACCGCACCCACCTTTAAACAYGGGGCTTGTAACTCAGCTCAC 1
|||||

RESULT 24
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:


```
; EARLIER FILING DATE: 1998-06-18
; EARLIER APPLICATION NUMBER: 60/050,244
; EARLIER FILING DATE: 1997-06-19
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Fast-SEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 3568
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (994)...(1863)
; NAME/KEY: misc feature
; LOCATION: (1)...(3568)
; OTHER INFORMATION: n = A,T,C or G
US-09-218-363-3

Query Match          5.4%; Score 34; DB 3; Length 3568;
Best Local Similarity 53.6%; Pred. No. 1.3;
Matches 67; Conservative 1; Mismatches 57; Indels 0; Gaps 0;

QY 511 TATAAACCCAGGATTCGAGCCGGCAACGGCAACCCCTTTGGGTCCCTCCCTTTGTAT 570
Db 650 TTTAAAAAAWATTTTTCGGSCCGGCGGTGTCACGCGCTGTAATCCAGGACTTTGGGA 591

QY 571 GGGCGCTCTGTTTCACTCTATTTCACTCTATTTAAATCTTTGCAACTGAAAAA 630
Db 590 GGGCGAGCCGAGATCAGCGCAATTTCACTCCAGCGCTGGCAACAGAGTGAATAATATCT 531

QY 631 AAAAA 635
Db 530 AAAAA 526

RESULT 28
US-09-621-976-9850
; Sequence 9850, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 9850
; LENGTH: 117
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-9850

Query Match          5.3%; Score 33.8; DB 4; Length 117;
Best Local Similarity 60.2%; Pred. No. 0.17;
Matches 56; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 543 ACCCCCTTTGGGTCCCTCCCTTTGTATGGCGCTCTGTTTCACTCTATTTCACTCTAT 602
Db 3 AACCTCTCAGGTCACCTTCCATGCTGTGTGGCTTTGTTCTTCCCTCTTTCGAATAAA 62

QY 603 TAAATCTTGCACCTGAAAAA 635
Db 63 TATTGCTGCTGTTCAAAAAA 95

RESULT 29
US-09-621-976-1047
; Sequence 1047, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
```

US-08-011-398B-1

```
Query Match      5.2%; Score 33; DB 1; Length 2417;
Best Local Similarity 58.8%; Pred. No. 2.1;
Matches 57; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY      539 GGCAACCCCTTTGGGTCCCTCCCTTTGTATGGGCGCTCTGTTTCACCTCTATTTCACCT 598
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      162 GGGAGGCCCTCTTGGAGACCTTACCCCTGGCTGTTTGGACTTTGTATACTTTAAATAATT 103

QY      599 CTATTAAATCTTCAACTGNAAAAAAAAAAAAAAAAAA 635
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      102 TAACTACCCTTAATTACTTAAAAAAAAAAAAAAAAAAAA 66
```

Search completed: January 22, 2005, 19:36:55
Job time : 61.7997 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM nucleic - nucleic search, using sw model

Run on: January 22, 2005, 12:12:40 ; Search time 344.872 Seconds
(without alignments)
10579.682 Million cell updates/sec

Title: US-09-319-156B-6
Perfect score: 635
Sequence: 1 cccgtatctttaacctctc.....tgaaaaaaaaaaaaaaaaaaaaa 635

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	635	100.0	635	8	US-08-979-847-102
2	635	100.0	635	15	US-10-114-104-102
3	617.4	97.2	2030	16	US-10-637-565-18
4	564.8	88.9	366710	18	US-10-719-993-6792
5	546.4	86.0	2074	16	US-10-416-642-4
6	545	85.8	1329	8	US-08-979-847-108
7	545	85.8	1329	15	US-10-114-104-108
8	543.4	85.6	21646	13	US-10-087-192-910
9	533.8	84.1	2946	13	US-10-114-893-134
10	533.8	84.1	2946	15	US-10-016-249-3
11	532.8	83.9	326014	9	US-09-731-231A-3
12	532.8	83.9	326014	17	US-10-751-985-3

13	525.8	82.8	1136	16	US-10-632-793-25	Sequence 25, Appl
14	522.8	82.3	283351	18	US-10-719-993-7065	Sequence 7065, Ap
15	522.8	82.3	1980090	18	US-10-719-993-6815	Sequence 6815, Ap
16	522.6	82.3	2782	16	US-10-632-793-26	Sequence 26, Appl
17	521	82.0	2782	16	US-10-632-793-61	Sequence 1, Appli
18	517.4	81.5	2930	9	US-09-902-535-1	Sequence 1, Appli
19	516.2	81.3	56093	10	US-09-873-367C-81	Sequence 81, Appl
20	511.4	80.5	7582	16	US-10-632-793-30	Sequence 30, Appl
21	510.6	80.4	161334	13	US-10-087-192-730	Sequence 730, App
22	500.8	78.9	285020	13	US-10-087-192-1666	Sequence 1666, Ap
23	499.6	78.0	1894	9	US-09-864-761-4444	Sequence 4444, Ap
24	495.6	78.0	3372	16	US-10-632-793-28	Sequence 28, Appl
25	493.2	77.7	2052	16	US-10-276-774-678	Sequence 678, App
26	482.6	76.0	22436	11	US-09-997-722-148	Sequence 148, App
27	481.4	75.8	849	16	US-10-220-120-15	Sequence 15, Appl
28	466	73.4	2349	15	US-10-104-047-1148	Sequence 1148, Ap
29	449	70.7	1564	15	US-10-204-887-10	Sequence 10, Appl
30	446.2	70.3	1393	16	US-10-220-120-17	Sequence 17, Appl
31	434.4	68.4	7974	14	US-10-198-846-9936	Sequence 9936, Ap
32	427.6	67.3	44063	17	US-10-322-281-718	Sequence 718, App
33	424.6	66.9	604	13	US-10-027-632-289680	Sequence 289680,
34	424.6	66.9	604	15	US-10-027-632-289680	Sequence 289680,
35	415.6	65.4	1394	17	US-10-466-531-47	Sequence 47, Appl
36	395.8	62.3	548	15	US-10-029-385-8554	Sequence 8554, Ap
37	388.4	61.2	23855	15	US-10-242-355-1175	Sequence 1175, Ap
38	383.4	60.4	657	13	US-10-027-632-224966	Sequence 224966,
39	383.4	60.4	657	15	US-10-027-632-224966	Sequence 224966,
40	373	58.7	439	9	US-09-764-881-183	Sequence 183, App
41	373	58.7	439	10	US-09-764-881-183	Sequence 183, App
42	373	58.7	439	16	US-10-242-747-183	Sequence 183, App
43	373	58.7	439	16	US-10-158-057-365	Sequence 365, App
44	371	58.4	587	13	US-10-027-632-321647	Sequence 321647,
45	371	58.4	587	15	US-10-027-632-321647	Sequence 321647,

ALIGNMENTS

RESULT 1

US-08-979-847-102

; Sequence 102, Application US/08979847

; Publication No. US20030039664A1

; GENERAL INFORMATION:

; APPLICANT: FERRON, HERVE

; APPLICANT: BESEME, FREDERIC

; APPLICANT: BEDIN, FREDERIC

; APPLICANT: PARANHOS-BACCALA, GLAUCIA

; APPLICANT: KOMURIAN-PRADEL, FLORENCE

; APPLICANT: JOLIVET-REYNAUD, COLETTE

; APPLICANT: MANDRAND, BERNARD

; APPLICANT: GARSON, JEREMY

; APPLICANT: TUKE, PHILIP

; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT

; TITLE OF INVENTION: THERAPEUTIC PURPOSES

; NUMBER OF SEQUENCES: 210

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OLIFF & BERRIDGE, PLC

; STREET: P.O. BOX 19928

; CITY: ALEXANDRIA

; STATE: VA

; COUNTRY: USA

; ZIP: 22320

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/979,847

; FILING DATE: 26-NOV-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 635 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-08-979-847-102

Query Match 100.0%; Score 635; DB 8; Length 635;
Best Local Similarity 100.0%; Pred. No. 1.5e-187; Indels 0; Gaps 0;
Matches 635; Conservative 0; Mismatches 0;

QY 1 CCTGTATCTTTAACTCTCTGTTAGTTTGTCTCTTCCAGAAATCAAACTGTAAACTA 60
Db 1 CCTGTATCTTTAACTCTCTGTTAGTTTGTCTCTTCCAGAAATCAAACTGTAAACTA 60
QY 61 CAAATGTTCTTCAATGGAGCCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
Db 61 CAAATGTTCTTCAATGGAGCCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
QY 121 CTGACCGGCTGCTAGCCCATCTCCGATGTTAATGACATTCGAAGCCACCCCTCCCGAG 180
Db 121 CTGACCGGCTGCTAGCCCATCTCCGATGTTAATGACATTCGAAGCCACCCCTCCCGAG 180
QY 181 GAAATCTCAACTGACAAACCCCTACTATGCCCCAAATTCAGCGGGAAGCAGTTAGACGGT 240
Db 181 GAAATCTCAACTGACAAACCCCTACTATGCCCCAAATTCAGCGGGAAGCAGTTAGACGGT 240
QY 241 CATCAGCCAACTCCCAACAGCACTTGGGTTTCTCTGTTGAGAGGGGAGCTGAGAGAC 300
Db 241 CATCAGCCAACTCCCAACAGCACTTGGGTTTCTCTGTTGAGAGGGGAGCTGAGAGAC 300
QY 301 AGGACTAGCTGGATTTCTTAGGCCAACGAAGATCCCTTAAGCCTTAGCTGGGAAGGTGACT 360
Db 301 AGGACTAGCTGGATTTCTTAGGCCAACGAAGATCCCTTAAGCCTTAGCTGGGAAGGTGACT 360
QY 361 GCATCCACCTCTAAACATGGGGCTTGCAACTAGCTCACAACCCGACCAATCAGAGAGCTC 420
Db 361 GCATCCACCTCTAAACATGGGGCTTGCAACTAGCTCACAACCCGACCAATCAGAGAGCTC 420
QY 421 ACTAAATGCTAATAGGCAAAATAGAGGTAAAGAAATAGCAATCATCTATTGCTG 480
Db 421 ACTAAATGCTAATAGGCAAAATAGAGGTAAAGAAATAGCAATCATCTATTGCTG 480
QY 481 AGAGCAGCGGGAGGACAGAGATCGGGATATAAACCAGGCAATTCGAGCCGGCAACGG 540
Db 481 AGAGCAGCGGGAGGACAGAGATCGGGATATAAACCAGGCAATTCGAGCCGGCAACGG 540
QY 541 CAACCCCTTTGGGTCCTCCCTTTGATGGGGCTCTGTTTTCACTATTTCACCTCT 600
Db 541 CAACCCCTTTGGGTCCTCCCTTTGATGGGGCTCTGTTTTCACTATTTCACCTCT 600
QY 601 ATTAATCTTGCACCTGAAAAAAGAAAAA 635
Db 601 ATTAATCTTGCACCTGAAAAAAGAAAAA 635

RESULT 2
US-10-114-104-102
Sequence 102, Application US/10114104
Publication No. US20030198647A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
BESEME, FREDERIC
BEDIN, FREDERIC
PARANHOS-BACCALA, GLAUCIA

KOMURIAN-PRADEL, FLORENCE
JOLIVET-REYNAUD, COLETTE
MANDRAND, BERNARD
GARSON, JEREMY
TUKER, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/114,104
FILING DATE: 03-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/979,847
FILING DATE: 26-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 635 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-10-114-104-102

Query Match 100.0%; Score 635; DB 15; Length 635;
Best Local Similarity 100.0%; Pred. No. 1.5e-187; Indels 0; Gaps 0;
Matches 635; Conservative 0; Mismatches 0;

QY 1 CCTGTATCTTTAACTCTCTGTTAGTTTGTCTCTTCCAGAAATCAAACTGTAAACTA 60
Db 1 CCTGTATCTTTAACTCTCTGTTAGTTTGTCTCTTCCAGAAATCAAACTGTAAACTA 60
QY 61 CAAATGTTCTTCAATGGAGCCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
Db 61 CAAATGTTCTTCAATGGAGCCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
QY 121 CTGACCGGCTGCTAGCCCATCTCCGATGTTAATGACATTCGAAGCCACCCCTCCCGAG 180
Db 121 CTGACCGGCTGCTAGCCCATCTCCGATGTTAATGACATTCGAAGCCACCCCTCCCGAG 180
QY 181 GAAATCTCAACTGACAAACCCCTACTATGCCCCAAATTCAGCGGGAAGCAGTTAGACGGT 240
Db 181 GAAATCTCAACTGACAAACCCCTACTATGCCCCAAATTCAGCGGGAAGCAGTTAGACGGT 240
QY 241 CATCAGCCAACTCCCAACAGCACTTGGGTTTCTGTTGAGAGGGGAGCTGAGAGAC 300
Db 241 CATCAGCCAACTCCCAACAGCACTTGGGTTTCTGTTGAGAGGGGAGCTGAGAGAC 300
QY 301 AGGACTAGCTGGATTTCTTAGGCCAACGAAGATCCCTTAAGCCTTAGCTGGGAAGGTGACT 360
Db 301 AGGACTAGCTGGATTTCTTAGGCCAACGAAGATCCCTTAAGCCTTAGCTGGGAAGGTGACT 360

361	Qy	GCATCCACCTCTAAACATGCGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC	420
361	Db	GCATCCACCTCTAAACATGCGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC	420
421	Qy	ACTAAAAATGCTAAATTAGGCAAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCGCTG	480
421	Db	ACTAAAAATGCTAAATTAGGCAAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCGCTG	480
481	Qy	AGAGCACAACGCGGGAGGACAAGGATCGGGATATAAACCAGGCATTCGAGCCGGCAACGG	540
481	Db	AGAGCACAACGCGGGAGGACAAGGATCGGGATATAAACCAGGCATTCGAGCCGGCAACGG	540
541	Qy	CAACCCCTTTGGGTGCCCTCCCTTTTGATGGCGCTCTGTTTTCACCTCTATTTCACCTCT	600
541	Db	CAACCCCTTTGGGTGCCCTCCCTTTTGATGGCGCTCTGTTTTCACCTCTATTTCACCTCT	600
601	Qy	ATTTAAATCTTGCACACTGAAAAAATAAAAAAAAAAAAAA	635
601	Db	ATTTAAATCTTGCACACTGAAAAAATAAAAAAAAAAAAAA	635

RESULT 3

```

US-10-637-565-18
; Sequence 18, Application US/10637565
; Publication No. US20040043381A1.
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: PERRON, Herve
; APPLICANT: KOMURIAN-PRADEL, Florence
; TITLE OF INVENTION: THE LTR REGION OF MSRV-1 AND THE PROTEINS IT ENCODES, AND PROBES
; TITLE OF INVENTION: METHODS FOR DETECTING MSRV-1 RETROVIRUS
;

```

Query Match	97.2%	Score 617.4;	DB 16;	Length 2030;
Best Local Similarity	98.3%;	Prod. No. 8.3e-182;		
Matches 624;	Conservative 0;	Mismatches 11;	Indels 0;	Gaps 0;

Qy	1	CCCTGTATCTTTAACTTAACTTGTCTCTCTCCAGATCAAACTGTAAACTA	60
Db	1396	CCCTGTATCTTTAACTTAACTTGTCTCTCTCCAGATGGAAGCTGTAAAGCTA	1455
Qy	61	CAAAATCTTCTTCAAATGGAGCACCATGGAGTCCATGACTAAGATCCACCGTGGACCC	120
Db	1456	CAGATGGTCTTACAAATGGAAACCCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC	1515
Qy	121	CTGACCGGCGCTGTAGCCCATGTCCTCGATGTTAATGACATTGAAGGACCCCTCCCGAG	180
Db	1516	CTGACCGGCGCTGTAGCCCATGTCCTCGATGTTAATGACATTGAAGGACCCCTCCCGAG	1575
Qy	181	GAATCTCAACTGCACACACCCCTACTATGCCCAATTCAGCGGGAACGATGTAGAGCGGT	240
Db	1576	GAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGGAACGATGTAGAGCGGT	1635
Qy	241	CATCAGCCAACTCTCCCAACAGCACTTTGGGTTTTCTCTTTGAGAGGGGGGACTGAGAGAC	300

Db	1636	CATCAGCCAACTCCCCCAA	CAGCACTTGGGTCTTCTGTTGAGAGGGGGGACTGAGAGAC	1695
Qy	301	AGGACTAGCTGGATTTCTT	CCTAGGCCAACGAGAATCCCTAAGCCCTAGCTCGGGAAGGTGACT	360
Db	1696	AGGACTAGCTGGATTTCTT	CCTAGGCCAACGAGAATCCCTAAGCCCTAGCTCGGGAAGGTGACT	1755
Qy	361	GCATCCACCTCTTAAACA	TATGCGGCTTGCACTTATAGCTCACACCCGACCAATCAGAGAGCTC	420
Db	1756	GCATCCACCTCTTAAACA	TATGCGGCTTGCACTTATAGCTCACACCCGACCAATCAGAGAGCTC	1815
Qy	421	ACTTAAATGCTAATTAGG	CGCAAAATAGGAGGTAAAGAAATAGCCCAATCATCTATTGCCTG	480
Db	1816	ACTTAAATGCTAATTAGG	CGCAAAATAGGAGGTAAAGAAATAGCCCAATCATCTATTGCCTG	1875
Qy	481	AGAGCAGCGGGAGGGACA	AGGATCGGGATATAAAACCCAGGCAATTCGAGCCGGCAACGG	540
Db	1876	AGAGCAGCGGGAGGGACA	AGGATCGGGATATAAAACCCAGGCAATTCGAGCCGGCAACGG	1935
Qy	541	CAACCCCTTTGGGTCCCCT	CTCTTTGTATGGCGCTCTGTTTTCACTCTATTTCACCTCT	600
Db	1936	CAACCCCTTTGGGTCCCCT	CTCTTTGTATGGCGCTCTGTTTTCACTCTATTTCACCTCT	1995
Qy	601	ATTAAATCTTTGCAACTG	AAAAAATAAAAAAAAAAAAAA	635
Db	1996	ATTAAATCTTTGCAACTG	AAAAAATAAAAAAAAAAAAAA	2030

RESULT 4

```

US-10-719-993-6792/c
; Sequence 6792, Application US/10719993
; Publication No. US20040285849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6792
; LENGTH: 366710
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(366710)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tab
US-10-719-993-6792

```

Query Match	88.9%;	Score 564.8;	DB 18;	Length 366710;
Best Local Similarity	93.4%;	Pred. 2.6e-164;		
Matches 590;	Conservative 0;	Mismatches 42;	Indels ~ 0;	Gaps 0;
Qy	1	CCCTGTATCTTTAACTCTTGTTTAAAGTTTGTCTTCCAGAAATCAAACTGTAAACTA	60	
Db	43899	CCCTGTATCTTTAAACCACTTGTTAAGTTTGTGTCTCCAGAAATTTAAACTGTAAACTA	43840	
Qy	61	CAAAATGTTCTTCAAAATGAGACACAGATGGAGTCCATGACTTAAGATCCACCGTGGACCC	120	
Db	43839	CAAAATGTTCTTCAAAATGAAACCCAGATGACATGTCATGTCTAAGATCTCCACGGACCC	43780	
Qy	121	CTGGACCGGCGCTGCTAGCGCACTCCGATGTTTAATGACATTGAAGGCACCCCTCCCGAG	180	
Db	43779	CTAGACTGCGCTGCTAGCCCCATGCTCTGGTGTTAATGACATCGAAGTCACCTCCTCTGAG	43720	
Qy	181	GAAATCTCAACTGCACAAACCCCTACTATGCCCCCAATTCAGCGGGAAGCAGTTAGACGGGT	240	
Db	43719	GAAATCTCAACTGCACAAACCCCTACTATGCCCCAGTTTCAGCAGGAAGCAGTTAGAGTGGT	43660	
Qy	241	CATCAGCGAACCTCCCCAAACGACACTTGGGTTTTCTGTTGAGAGGGGGGACTCAGAGAC	300	
Db	43659	CATCAGCGAACCTCCACCAACAGCACTTGGGTTTTCTGTTGAGAGAGGGGACTCAGAGAC	43600	

QY	301	AGGACTAGCTGGATTTCCTAGGCCAAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGACT	360
Db	43599	AGGACTAGCTGGATTTCCTAGGCCGATTAAAGAATCCCTAAACCTAGCTGGGAAGGTGACC	43540
QY	361	GCATCCACCTCTTAACATCGGGCTTGCACCTTAGCTCACACCCGACCAATCAGAGAGCTC	420
Db	43539	GCATCCACCTTTAAACACGGGCTTGCACCTTAGCTCACACCAACCAATCAGAGAGCTC	43480
QY	421	ACTAAATGCTAAATTAGGCAAAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGGCTG	480
Db	43479	ACTAAATGCTAAATTAGGCAAAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGGCTG	43420
QY	481	AGAGCACAGGGAGGGAACAAGATCGGATATAAACCCAGGCATTCGAGCGGCAACGG	540
Db	43419	AGAGCACAGTGGGAGGGAACAAGATTGCAATAATAAACCCAGGCATTCGAGCGGCAACGG	43360
QY	541	CAACCCCTTTGGTCCCTCCCTTTGTATGGCGCTCTGTTTTCACTCTATTTCACTCT	600
Db	43359	CAACCGCTTTGGTCCCTTCCTTTGTATGGAGCTCTGTTTTCACTCTATTTCACTCT	43300
QY	601	ATTAAATCTTTGCAACTGAAAAAATAAAAAA	632
Db	43299	ATTAAATCTTTGCAACTGAAAAAATAAAAAA	43268

```

RESULT 5
US-10-416-642--4
; Sequence 4, Application US/10416642
; Publication No. US20040043452A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: ARVIZU, Chandra
; TITLE OF INVENTION: EMBRYOGENESIS ASSOCIATED PROTEINS
; FILE REFERENCE: PP-0842 PCT
; CURRENT APPLICATION NUMBER: US/10/416,642
; CURRENT FILING DATE: 2003-05-13
; PRIOR APPLICATION NUMBER: 60/249,407
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 2074
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040043452A1 7477736CB1
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 1995
; OTHER INFORMATION: a, t, c, g, or other
US-10-416-642--4

```

1639	GAATCTCAACTCGACAAACCCCTACTATATGCCCAATTCAGCAGGAAGCAGTTAGACGGCT	1699
241	CATCAGCCAACTCCCAACAGCAGCACTTGGGTTTTCTGTGTGAGAGGGGACCTGAGAGAC	300
1699	CGTCAGTCAACCTCCCAACAGCAGCACTTGGGTTTTCTGTGTGAGAGGGGACCTGAGAGAC	1758
301	AGACCTAGCTGGATTTCTTAGGCCAAGAAAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT	360
1759	AGGACTAGCTGGATTTCTTAGGCCGGAATTAAGAAATCCCTAAGCCTAGCTGGGAAGGTGACC	1818
361	GCATCCACCTCTAAACATCGGGCTTCCAACTTAGCTTCACACCCGACCAATCAGAGAGCTC	420
1819	GGCTCACCTTTAAACACGGGGTTCGAACTTAGCTTCACACCCACCAATCAGAGAGCTC	1878
421	ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCCTG	480
1879	ACTAAATGCTAATTAGGCAAAACAGGAGGTAAAGAAATAGCCAATCATCTATTGCCCTG	1938
481	AGAGCAGCGGGAGGACAAGATCGGGATATAAACCCAGGATTCGAGCCGGCAACGG	540
1939	AGAGCAGTGGGAGGACAAGATTCGAAATATAAACCCAGGATTCGAGCCAGC-ANAG	1997
541	CAACCCCTTTGGTCCCTCCCTTGTATGGCGCTCTGTTTTCACTCTATTTCCTCT	600
1998	CAACCCCTTTGGTCCCTTCCCTTGTATGGGAGCTCTGTTTTCACTCTATTTCCTCT	2057
601	ATTAATCTTGCAACTG	617
2058	ATTAATCTTGCAACTG	2074

RESULT 6

US-08-979-108

; Sequence 108, Application US/08979847

; Publication No. US20030039664A1

; GENERAL INFORMATION:

; APPLICANT: PERRON, HERVE

; APPLICANT: BESEME, FREDERIC

; APPLICANT: BEDIN, FREDERIC

; APPLICANT: PARANHOS-BACCALA, GLAUCIA

; APPLICANT: KOMURIAN-PRADEL, FLORENCE

; APPLICANT: JOLIVET-REYNAUD, COLETTE

; APPLICANT: MANDRAND, BERNARD

; APPLICANT: GARSON, JEREMY

; APPLICANT: TUKE, PHILIP

; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, P

; TITLE OF INVENTION: THERAPEUTIC PURPOSES

; NUMBER OF SEQUENCES: 210

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OLIFF & BERRIDGE, PLC

; STREET: P.O. BOX 19928

; CITY: ALEXANDRIA

; STATE: VA

; COUNTRY: USA

; ZIP: 22320

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/979,847

; FILING DATE: 26-NOV-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: BERRIDGE, WILLIAM P.

; REGISTRATION NUMBER: 30,024

; REFERENCE/DOCKET NUMBER: WPB 39046A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-836-6400

; TELEFAX: 703-836-2787

; INFORMATION FOR SEQ ID NO: 108:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1329 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-979-847-108

Query Match 85.8%; Score 545; DB 8; Length 1329;
Best Local Similarity 92.4%; Pred. No. 2.8e-159;
Matches 572; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 1 CCTGTATCTTTAACTCCTCTTGTAGTTGTTCTTCCAGAACTCAAACTGTAAGACTA 60
DB 257 CCTGTATCTTCAACTCTCTTGTAGTTGTTCTTCCAGAACTGAACTGTAAGACTA 316

QY 61 CAAATTTGTTCTTCAATGGAGCACCAGATGAGTCCATGACTAAGATCCACCGTGGACCC 120
DB 317 CAATAGTTCTTCAATGGAAACCCAGATGAGTCCATGACTAATCTACCGTGGACCC 376

QY 121 CTGGACCGGCTGTGACCCATGCTCGATGTTAATGACATTTGAAGGACCCCTCCCGAG 180
DB 377 CTGGACCGGCTGTGACCTATGCTCTGATGTTAATGACATTTGAAGTCAACCCCTCCCGAG 436

QY 181 GAAATCTCACTGCACAAACCCCTACTATGCCCAATTCAGCGGGAACGAGTTAGAGCGGT 240
DB 437 GAAATCTCACTGCACAAACCCCTACTATGCCCAATTCAGTGGAAAGCTTAGAGCGGT 496

QY 241 CATCAGCAACCTCCCCAACAGACACTTGGGTTTTCTGTTGAGAGGGGAGCTGAGAGAC 300
DB 497 TGTGACGCAACCTCCCCAACAGACTTGGGTTTTCTGTTGAGAGGGTGGAGCTGAGAGAC 556

QY 301 AGACTAGCTGGAATTTCTTAGGCCAACAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
DB 557 AGACTAGCTGGAATTTCTTAGGCTGACTAAGAAATCCCNAAAGCTTANCTGGGAAGGTGACC 616

QY 361 GCATCCACCTTAAACATAGGGGCTTGCACCTAGCTCAGACCCGACCAATCAGAGAGCTC 420
DB 617 GCATCCACCTTAAACATAGGGGCTTGCACCTAGCTCAGACCCGACCAATCAGAGAGCTC 676

QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 480
DB 677 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 736

QY 481 AGAGCAGCGGAGGAGCAAGGATCGGATATAAACCCAGGCAITTCGAGCGGCAACGG 540
DB 737 AGAGCAGCGGAGGAGCAAGGATCGGATATAAACCTCAGGCAITTCAGSCAGCAACAG 796

QY 541 CAACCCCTTTGGGTCCTCCCTCTTGTATGGGGCTCTGTTTTCAGTCTATTTCAGTCT 600
DB 797 CAACCCCTTTGGGTCCTCCCTCTTGTATGGGAGCTCTGTTTTCAGTCTATTTCAGTCT 856

QY 601 ATTAATCTTGCAACTGAA 619
DB 857 ATTAATCTGCAACTGCA 875

RESULT 7

US-10-114-104-108
; Sequence 108, Application US/10114104
; Publication No. US20030198647A1
; GENERAL INFORMATION:
; APPLICANT: PERON, HERVE
; BESEME, FREDERIC
; BEDIN, FREDERIC
; PARANHOS-BACCALA, GLAUCIA
; KOMURIAN-PRADEL, FLORENCE
; JOLIVET-RENAUD, COLETTE
; MANDRAND, BERNARD
; GARSON, JEREMY
; TUBE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL

; THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patencin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114.104
; FILING DATE: 03-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/979,847
; FILING DATE: 26-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BERRIDGE, WILLIAM P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 39046A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 108:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1329 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; SEQUENCE DESCRIPTION: SEQ ID NO: 108:
; US-10-114-104-108

Query Match 85.8%; Score 545; DB 15; Length 1329;
Best Local Similarity 92.4%; Pred. No. 2.8e-159;
Matches 572; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 1 CCTGTATCTTTAACTCCTCTTGTAGTTGTTCTTCCAGAACTCAAACTGTAAGACTA 60
DB 257 CCTGTATCTTCAACTCTCTTGTAGTTGTTCTTCCAGAACTGTAAGACTA 316

QY 61 CAAATTTGTTCTTCAATGGAGCACCAGATGAGTCCATGACTAAGATCCACCGTGGACCC 120
DB 317 CAATAGTTCTTCAATGGAAACCCAGATGAGTCCATGACTAATCTACCGTGGACCC 376

QY 121 CTGGACCGGCTGTGACCCATGCTCCGATGTTAATGACATTTGAAGGCAACCCCTCCCGAG 180
DB 377 CTGGACCGGCTGTGACTATGCTCTGATGTTAATGACATTTGAAGTCAACCCCTCCCGAG 436

QY 181 GAAATCTCACTGCACAAACCCCTACTATGCCCAATTCAGCGGGAACGAGTTAGAGCGGT 240
DB 437 GAAATCTCACTGCACAAACCCCTACTATGCCCAATTCAGTGGAAAGCTTAGAGCGGT 496

QY 241 CATCAGCAACCTCCCCAACAGACACTTGGGTTTTCTGTTGAGAGGGGAGCTGAGAGAC 300
DB 497 TGTGACGCAACCTCCCCAACAGACTTGGGTTTTCTGTTGAGAGGGTGGAGCTGAGAGAC 556

QY 301 AGACTAGCTGGAATTTCTTAGGCCAACAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
DB 557 AGACTAGCTGGAATTTCTTAGGCTGACTAAGAAATCCCNAAAGCTTANCTGGGAAGGTGACC 616

QY 361 GCATCCACCTTAAACATAGGGGCTTGCACCTAGCTCAGACCCGACCAATCAGAGAGCTC 420
DB 617 GCATCCACCTTAAACATAGGGGCTTGCACCTAGCTCAGACCCGACCAATCAGAGAGCTC 676

QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 480
DB 677 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 736

```
QY 481 AGAGCAGCGGAGGAGCAAGATCGGATATAAACCAGGCAATTCGAGCCGCGCAACGG 540
Db 737 AGAGCAGCGGAGGAGCAAGATCGGATATAAACCAGGCAATTCGAGCCGCGCAACAG 796
QY 541 CAACCCCTTTGGGTCCCTCCCTTTGTATGGCGCTCTGTTTCACTCTTCACTCT 600
Db 797 CAACCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTCACTCTTCACTCT 856
QY 601 ATTAATCTTGCNACTGAA 619
Db 857 ATTAATCATGCAACTGCA 875

RESULT 8
US-10-087-192-910
; Sequence 910, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 910
; LENGTH: 21646
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc (21646)
; LOCATION: (1)..(21646)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-910

Query Match 85.6%; Score 543.4; DB 13; Length 21646;
Best Local Similarity 94.8%; Pred. No. 3.3e-158;
Matches 562; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1 CCCTGTATCTTTAACTCTTGTAACTTTGTCTTCCAGATCAAACTGTAAACTA 60
Db 19571 CCCTGTATCTTTAACTCTTGTAACTTTGTCTTCCAGATTTGAACTGTAAACTA 19630
QY 61 CAAATTTCTTCAAATGGAGCACAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
Db 19631 CAAATGGTTCTTCAAATGGAGCCCCAGATGCAGTCCATGACTAAGATCCACAGACCC 19690
QY 121 CTGACCGGCTGCTAGCCCATGCTCCGATGTTAATGACATTAAGAGCAACCCCTCCGAG 180
Db 19691 CTGACCGAGCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCAACCCCTCCGAG 19750
QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAGCAAGCTTAGAGCGGT 240
Db 19751 GAAATCTCAACTGTGCAACCCCTACTATGCCCAATTCAGCGGAGCAAGCTTAGAGCGGT 19810
QY 241 CATCAGCAACCTCCCAACAGCACTTGGGTTTTCTGTTGAGAGGGGGGACTGAGAGAC 300
Db 19811 CATCAGCAACCTCCCAACAGCACTTGGGTTTTCTGTTGAGAGGGGGGACTGAGAGAC 19870
QY 301 AGGACTAGCTGGATTTCTTAGGCCAAACGAAATCCCTAAGCTTAGCTGGGAAGGTGACT 360
Db 19871 AAGACTAGCTGGATTTCTTAGGCCAACTAAGAAATCCCTAAGCTTAGCTGGGAAGGTGACT 19930
QY 361 GCATCCACCTTAAACATGGGGCTTGCAACTTACTCACCCTGACCAATCAGAGAGCTC 420
Db 19931 GCATCCACCTTTAAACATGGGGCTTGCAACTTACTCACCCTGACCAATCAGAGAGCTC 19990
```

```
QY 421 ACTAAATGCTAATTTAGGCAAAATAGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 480
Db 19991 ACTAAATGCTAATTTAGGCAAAATAGAGGTAAAGAAATAGCCAATCATCTATTGCCTG 20050
QY 481 AGAGCAGCGGAGGAGCAAGATCGGATATAAACCAGGCAATTCGAGCCGCGCAACGG 540
Db 20051 AGAGCAGCGGAGGAGCAAGATCGGATATAAACCAGGCAATTCGAGCCGCGCAATGG 20110
QY 541 CAACCCCTTTGGGTCCCTCCCTTTGTATGGCGCTCTGTTTCACTCTATT 593
Db 20111 CAACCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTCACTCTATT 20163

RESULT 9
US-10-114-893-134
; Sequence 134, Application US/10114893
; Publication No. US20020193567A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Carlin-Duckett, McKeough
; APPLICANT: Kelleher, Kerry S.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: GI 6000-10A
; CURRENT APPLICATION NUMBER: US/10/114,893
; CURRENT FILING DATE: 2002-04-02
; EARLIER APPLICATION NUMBER: 09/413,232
; EARLIER FILING DATE: 1999-10-06
; NUMBER OF SEQ ID NOS: 321
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 134
; LENGTH: 2946
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-114-893-134

Query Match 84.1%; Score 533.8; DB 13; Length 2946;
Best Local Similarity 91.5%; Pred. No. 1.3e-155;
Matches 581; Conservative 0; Mismatches 42; Indels 12; Gaps 1;

QY 1 CCCTGTATCTTTAACTCTTGTAACTTTGTCTTCCAGATCAAACTGTAAACTA 60
Db 2323 CCCTGTATCTTTAACTCTTGTAACTTTGTCTTCCAGATCGAAGCTGTAAACTA 2382
QY 61 CAAATTTCTTCAAATGGAGCACAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
Db 2383 CAAATGGAGCCCCAGATGCACTAAGATCTACCCAGACCC 2430
QY 121 CTGACCGGCTGCTAGCCCATGCTCCGATGTTAATGACATTAAGAGCAACCCCTCCGAG 180
Db 2431 CTGACCGGCTGCTAGCCCATGCTCCGATGTTAATGACATCAAGAGCAACCCCTCTGAG 2490
QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAGCAAGCTTAGAGCGGT 240
Db 2491 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAGCAAGCTTAGAGCGGT 2550
QY 241 CATCAGCAACCTCCCAACAGCACTTGGGTTTTCTGTTGAGAGGGGGGACTGAGAGAC 300
Db 2551 CAGCGCAACCTCCCAACAGCACTTGGGTTTTCTGTTGAGAGGGGGGACTGAGAGAC 2610
QY 301 AGGACTAGCTGGATTTCTTAGGCCAAACGAAATCCCTAAGCTTAGCTGGGAAGGTGACT 360
Db 2611 AGGACTAGCTGGATTTCTTAGGCCAACTAAGAAATCCCTAAGCTTAGCTGGGAAGGTGACT 2670
```

QY 361 GCATCCACCTTAACATGGGCTTGCACTTAGCTCACACCCGACCAATCAGAGCTC 420
DB 2671 ACATCCACCTTTAAACACGGGGCTTGCACTTAGCTCACACCTGACCAATCAGAGCTC 2730
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGTAAAGAAATAGCCAAATCATCTATTGGCTG 480
DB 2731 ACTAAATGCTAATTAGGCAAAATAGGAGTAAAGAAATAGCCAAATCATCTATTGGCTG 2790
QY 481 AGAGCAGCGGGAGGACCAAGGATCGGATATAAACCAGGCAATTCGAGCCGGCAACGG 540
DB 2791 AGAGCAGCGGGAGGACCAAGGATCGGATATAAACCAGGCAATTCGAGCCGGCAACGG 2850
QY 541 CAACCCCTTTGGGTCCTCCCTTTGATGGGCTCTCTTTTCACTCTATTTCACCTCT 600
DB 2851 CAACCCCTTTGGGTCCTCCCTTTGATGGGCTCTCTTTTCACTCTATTTCACCTCT 2910
QY 601 ATTAAATCTTGCAACTGAAAAAAGAAAAA 635
DB 2911 ATTAAATCTTGCAACTGAAAAAAGAAAAA 2945

RESULT 10

US-10-016-249-3

; Sequence 3, Application US/10016249
; Publication No. US20030100053A1

; GENERAL INFORMATION:

; APPLICANT: Jacobs, Kenneth

; APPLICANT: McCoy, John M.

; APPLICANT: Lavallie, Edward R.

; APPLICANT: Collins-Racie, Lisa A.

; APPLICANT: Evans, Cheryl

; APPLICANT: Merberg, David

; APPLICANT: Treacy, Maurice

; APPLICANT: Mi, Sha

; APPLICANT: Genetics Institute, Inc.

; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

; FILE REFERENCE: 6006B-Jul72A

; CURRENT APPLICATION NUMBER: US/10/016,249

; CURRENT FILING DATE: 2001-10-30

; PRIOR FILING DATE: 1998-10-20

; NUMBER OF SEQ ID NOS: 62

; SOFTWARE: PatentIn ver. 2.0

; SEQ ID NO 3

; LENGTH: 2946

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-016-249-3

Query Match 84.1%; Score 533.8; DB 15; Length 2946;

Best Local Similarity 91.5%; Pred. No. 1.3e-155;

Matches 581; Conservative 0; Mismatches 42; Indels 12; Gaps 1;

QY 1 CCCTGTATCTTTAACTCCTTGTTAGTTTGTCTCTCCAGAAATCAAACTGTAATACTA 60
DB 2323 CCCTGTATCTTTAACTCCTTGTTAGTTTGTCTCTCTCCAGAAATCGAAGCTGTAATACTA 2382
QY 61 CAAATTGTTCTTCAATGGAGACCCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
DB 2383 -----CAATGGAGACCCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 2430
QY 121 CTGACCGGCTGCTAGCCCATGCTCCGATGTTAATGACATTTGAAGGACCCCTCCCGAG 180
DB 2431 CTGACCGGCTGCTAGCCCATGCTCCGATGTTAATGACATCAAGGACCCCTCCCGAG 2490
QY 181 GAAATCTCACTGACAAACCTCTACTATGCCCAATTCAGGGGAGCAGTTAGAGCGGT 240
DB 2491 GAAATCTCAGTGCACAACTCTACTACGCCCAATTCAGGGGAGCAGTTAGAGCGGT 2550
QY 241 CATCAGCAACCTCCCAACAGCACTTGGGTTTTTCTGTTGAGAGGGGGGACTGAGAGAC 300
DB 2551 CGTGGCCAACTCCCAACAGCACTTGGGTTTTTCTGTTGAGATGGGGGACTGAGAGAC 2610

QY 301 AGGACTAGCTGGATTTCTTAGGCCAACGAGAAATCCCTAGCCTAGCTGCGAAGGTGACT 360
DB 2611 AGGACTAGCTGGATTTCTTAGGCCAACGAGAAATCCCTAGCCTAGCTGCGAAGGTGACT 2670
QY 361 GCATCCACCTTAACATGGGCTTGCACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
DB 2671 ACATCCACCTTTAAACACGGGGCTTGCACTTAGCTCACACCTGACCAATCAGAGAGCTC 2730
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGTAAAGAAATAGCCAAATCATCTATTGGCTG 480
DB 2731 ACTAAATGCTAATTAGGCAAAATAGGAGTAAAGAAATAGCCAAATCATCTATTGGCTG 2790
QY 481 AGAGCAGCGGGAGGACCAAGGATCGGATATAAACCAGGCAATTCGAGCCGGCAACGG 540
DB 2791 AGAGCAGCGGGAGGACCAAGGATCGGATATAAACCAGGCAATTCGAGCCGGCAACGG 2850
QY 541 CAACCCCTTTGGGTCCTCCCTTTGATGGGCTCTCTTTTCACTCTATTTCACCTCT 600
DB 2851 CAACCCCTTTGGGTCCTCCCTTTGATGGGCTCTCTTTTCACTCTATTTCACCTCT 2910
QY 601 ATTAAATCTTGCAACTGAAAAAAGAAAAA 635
DB 2911 ATTAAATCTTGCAACTGAAAAAAGAAAAA 2945

RESULT 11

US-09-731-231A-3/c

; Sequence 3, Application US/09731231A

; Patent No. US20020082189A1

; GENERAL INFORMATION:

; APPLICANT: GUGLER, Karl et al

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: CL001007

; CURRENT APPLICATION NUMBER: US/09/731,231A

; CURRENT FILING DATE: 2000-12-07

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 326014

; TYPE: DNA

; ORGANISM: Human

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)...(326014)

; OTHER INFORMATION: n = A, T, C or G

US-09-731-231A-3

Query Match 83.9%; Score 532.8; DB 9; Length 326014;

Best Local Similarity 90.5%; Pred. No. 2.4e-154;

Matches 583; Conservative 0; Mismatches 52; Indels 9; Gaps 1;

QY 1 CCCTGTATCTTTAACTCCTTGTTAGTTTGTCTCTCCAGAAATCAAACTGTAATACTA 60
DB 170822 CCCTGTATCTTTAACTCCTTGTTAGTTTGTCTCTCTCCAGAAATCGAAGCTGTAATACTA 170763
QY 61 CAAATTGTTCTTCAATGGAGACCCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
DB 170762 CAAATTGTTCTTCAATGGAGACCCAGATGGAGTCCATGACTAAGATCTACCGAGACCC 170703
QY 121 CTGACCGGCTGCTAGCCCATGCTCCGATGTTAATGACATTTGAAGGACCCCTCCCGAG 180
DB 170702 CTGACCGGCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGACCCCTCCCGAG 170643
QY 181 GAAATCTCACTGACAAACCTCTACTATGCCCAATTCAGGGGAGCAGTTAGAGCGGT 240
DB 170642 GAAATCTCACTGACAAACCTCTACTACGCCCAATTCAGGGGAGCAGTTAGAGCGGT 170583
QY 241 CATCAGCAACCTCCCAACAGCACTTGGGTTTTTCTGTTGAGAGGGGGGACTGAGAGAC 300
DB 170582 CATCAGCAACCTCCCAACAGCACTTGGGTTTTTCTGTTGAGATGGGGGACTGAGAGAC 170523


```
Qy 327 CGAAGATCCCTAAGCTAGCTGGGAGTGTGCTCATCCACTTAAACATGGGGCTTG 386
Db 1777383 CTAAGAAATCCCTAAGCTAGCTGGGAGTGTGCTCATCCACTTAAACATGGGGCTTG 1777442
Qy 387 CAACTTAGCTCACACCCGACCAATCAGAGAGTCTAATAATGCTAATTAGGCAAAAATA 446
Db 1777443 CAACTTAGCTCACAC---CCACACAGAGAGTCTAATAATGCTAATTAGGCAAAAATA 1777498
Qy 447 GAGGTAAAGAAATAGCAATCACTATTGCTGTAGAGACACAGCGGGAGGCAAGGATC 506
Db 1777499 GAGGTACAGAAATAGCAATCACTATTGCTGTAGAGACACAGCGGGAGGCAAGGATC 1777558
Qy 507 GGGATATAAACCCAGGCAATTCGAGCCGGCAACCGCAACCCCTTTGGGTCCCTCCCTTT 566
Db 1777559 AGGATATAAACCCAGGCAATTCGAGCTGGCAACCGCAACCCCTTTGGGTCCCTCCCTTT 1777618
Qy 567 GTATGGGCGCTGTGTTTCTACTCTATTCTACTCTATTAAATCTTGCAACTGCAAAAAAATA 626
Db 1777619 GTATGGGAGCTGTGTTTCTCA--CTGTTTCACTCTATTAAATCTTGCAACTGCAAAAAAATA 1777676
Qy 627 AAAAAAAA 634
Db 1777677 AAAAAAAA 1777684
```

```
RESULT 16
US-10-632-793-26
; Sequence 26, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: MALLET, Francois
; APPLICANT: VOISSET, Cecile
; TITLE OF INVENTION: ENDOGENOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
; TITLE OF INVENTION: AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
; FILE REFERENCE: 110048
; CURRENT APPLICATION NUMBER: US/10/632,793
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: US/09/869,927
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: PCT/FR00/00144
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: FR 99/00888
; PRIOR FILING DATE: 1999-01-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
; LENGTH: 2782
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-632-793-26
```

```
Query Match 82.3%; Score 522.6; DB 16; Length 2782;
Best Local Similarity 90.4%; Pred. No. 4e-152;
Matches 574; Conservative 0; Mismatches 49; Indels 12; Gaps 1;

Qy 1 CCTGTATCTTTAACTCTCTGTTAAGTTGTCTCTCCAGAAATCAAAACTGTAAACTA 60
Db 2158 CCTGTATCTTTGACCTCTCTGTTAAGTTGTCTCTCCAGAAATCGAAGTGTGAACATA 2217
Qy 61 CAAATTTGTTCTTCAATGGAGCACCAAGATGGAGTCCATGACTAAGATCCACCGTGACCC 120
Db 2218 -----CAAATGGAGCCCAAGATGGAGTCCATGACTAAGATCCACCGTGACCC 2265
Qy 121 CTGACCGGCTGTAGCCCATGCTCGATGTTAATGACATTAAGGACCCCTCCCGAG 180
Qy 2266 CTGACCGGCTGTAGCCCATGCTCGATGTTAATGACATTAAGGACCCCTCCCGAG 2325
Qy 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGGAAGAGCTTAGAGCGGT 240
Db 2326 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGGAAGAGCTTAGAGCGGT 2385
```

```
Qy 241 CATCAGCCACCTCCCAACAGCACTGGGTTTTCTGTTGAGAGGGGAGCTGAGAGAC 300
Db 2386 GGTGGCCAACTCCCAACAGCACTTAGGTTTTCTGTTGAGATGGGAGCTGAGAGAC 2445
Qy 301 AGGACTAGCTGGATTTCTTAGGCCCAACGAGAATCCTTAAGCCTAGCTGGGAAGGTGACT 360
Db 2446 AGGACTAGCTGGATTTCTTAGGCCCAACGAGAATCCTTAAGCCTAGCTGGGAAGGTGACT 2505
Qy 361 GCATCCACCTCTAAACATAGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
Db 2506 ACATCCACCTTTAAACAGCGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 2565
Qy 421 ACTAAATGCTAATAGGCAAAAATAGAGGTAAAGAAATAGCCCAATCATCTATTGCTGT 480
Db 2566 ACTAAATGCTAATAGGCAAAAATAGAGGTAAAGAAATAGCCCAATCATCTATTGCTGT 2625
Qy 481 AGAGCACAGCGGGAGGAGCAAGGATCGGATATAAACCCAGGCAATTCGAGCCGCAACGG 540
Db 2626 AGAGCACAGCGGGAGGAGCAAGGATCGGATATAAACCCAGGCAATTCGAGCCGCAACGG 2685
Qy 541 CAAACCCCTTTGGGTCCCTCCCTTTGTATGGGCGCTCTGTTTCACTCTATTTCACCTCT 600
Db 2686 CAAACCCCTTTGGGTCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACCTCT 2745
Qy 601 ATTAATCTTGCAACTGAAAAAATAAAAAA 635
Db 2746 ATTAATCTTGCAACTGAAAAAATAAAAAA 2780
```

```
RESULT 17
US-10-133-036-1
; Sequence 1, Application US/10133036
; Publication No. US20040054133A1
; GENERAL INFORMATION:
; APPLICANT: Conrad, Bernard
; APPLICANT: Mach, Bernard
; TITLE OF INVENTION: Multiple Sclerosis-Related Superantigen
; FILE REFERENCE: 23135-507
; CURRENT APPLICATION NUMBER: US/10/133,036
; CURRENT FILING DATE: 2002-04-26
; PRIOR APPLICATION NUMBER: PCT/EP00/10659
; PRIOR FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2782
; TYPE: DNA
; ORGANISM: Human endogenous retrovirus
US-10-133-036-1
```

```
Query Match 82.0%; Score 521; DB 16; Length 2782;
Best Local Similarity 90.2%; Pred. No. 1.3e-151;
Matches 573; Conservative 0; Mismatches 50; Indels 12; Gaps 1;

Qy 1 CCTGTATCTTTAACTCTCTGTTAAGTTGTCTCTTCGAAATCAAAACTGTAAACTA 60
Db 2158 CCTGTATCTTTGACCTCTCTGTTAAGTTGTCTCTTCGAAATCGAAGTGTGAACATA 2217
Qy 61 CAAATTTGTTCTTCAATGGAGCACCAAGATGGAGTCCATGACTAAGATCCACCGTGACCC 120
Db 2218 -----CAAATGGAGCCCAAGATGGAGTCCATGACTAAGATCCACCGTGACCC 2265
Qy 121 CTGACCGGCTGTAGCCCATGCTCGATGTTAATGACATTAAGGACCCCTCCCGAG 180
Db 2266 CTGACCGGCTGTAGCCCATGCTCGATGTTAATGACATTAAGGACCCCTCCCGAG 2325
Qy 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGGGGAGCAGTTAGAGCGGT 240
Db 2326 GAAATCTCAGCTGCACAAACCTCTACTACGCCCAATTCAGCGGAGCAGTTAGAGCGGT 2385
Qy 241 CATCAGCCAACTCCCAACAGCACTTGGGTTTTCTGTTGAGAGGGGGAAGCTGAGAGAC 300
Db 2386 GGTGGCCAACTCCCAACAGCACTTAGGTTTTCTGTTGAGATGGGAGCTGAGAGAC 2445
```

```
QY 301 AGGACTAGCTGGATTTCTTAGGCGCAACGAAGATCCCTTAAGCCTAGCTAGTGGGAAGGTGACT 360
Db 2446 AGGACTAGCTGGATTTCTTAGGCGCTAGCTAAGATCCCTTAAGCCTAGCTAGTGGGAAGGTGACC 2505
QY 361 GCATCCACCTCTAAACATGCGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
Db 2506 ACATCCACCTTTAAACAGCGGGCTTGCACTTAGCTCACACCTGACCAATCAGAGAGCTC 2565
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 480
Db 2566 ACTAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 2625
QY 481 AGAGCACAGCGGAGGACAGATCGGGATATAAACCAGGCAATTCGAGCCGCGCAACGG 540
Db 2626 AGAGCACAGCAGGAGGACCAATGATCGGGATATAAACCAGGCAATTCGAGCCGCGCAACGG 2685
QY 541 CAACCCCTTTGGGTCCCTCCCTTTGTATGGGGCTCTGTTTCACTCTATTTCACCTCT 600
Db 2686 CAACCCCTTTGGGTCCCTCCCTTTGTATGGGGCTCTGTTTCACTCTATTTCACCTCT 2745
QY 601 ATTAATCTTGCACTGAAAAAAGAAAAA 635
Db 2746 ATTAATCTTGCACTGAAAAAAGAAAAA 2780
```

RESULT 18

US-09-902-535-1

; Sequence 1, Application US/09902535

; Patent No. US20020102530A1

; GENERAL INFORMATION:

; APPLICANT: Keith, Jr., James C.

; APPLICANT: McCoy, John M.

; APPLICANT: Mi, Sha

; TITLE OF INVENTION: Methods and compositions for diagnosing

; and treating preeclampsia and gestational trophoblast

; disorders

; FILE REFERENCE: GIN-600884

; CURRENT APPLICATION NUMBER: US/09/902,535

; CURRENT FILING DATE: 2001-07-09

; PRIOR APPLICATION NUMBER: 60/216,657

; PRIOR FILING DATE: 2000-07-06

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 2930

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (930)...(2546)

US-09-902-535-1

Query Match 81.5%; Score 517.4; DB 9; Length 2930;

Best Local Similarity 91.4%; Pred. No. 1.7e-150;

Matches 564; Conservative 0; Mismatches 41; Indels 12; Gaps 1;

QY 1 CCCTGTATCTTTAACTCTCTTTGTTAACTTTGTCTCTCCAGAAATCAAACTGTAATAACTA 60

Db 2325 CCCTGTATCTTTAACTCTCTTTGTTAACTTTGTCTCTCCAGAAATCGAGCTGTAAACTA 2384

QY 61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGATCAAGATCCACCGTGGACCC 120

Db 2385 -----CAAATGGAGCACCAGATGGAGTCCAGATCAAGATCTACCGCAGACCC 2432

QY 121 CTGACCGGCTGTACCGCATCTCGATGTTTAATGACATGAGGACCCCTCCCGAG 180

Db 2433 CTGACCGGCTGTACCGCATCTCGATGTTTAATGACATCAAAAGGACCCCTCCCTGAG 2492

QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAAGCAGTTAGAGCGGT 240

Db 2493 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGAAGCAGTTAGAGCGGT 2552

```
QY 241 CATCAGCAACCTCCCAACAGCACTTGGGTTTTCTGTTGAGGGGGGACTGAGAGAC 300
Db 2553 CGTCGGCACAACCTCCCAACAGCACTTGGGTTTTCTGTTGAGATGGGGGACTGAGAGAC 2612
QY 301 AGGACTAGCTGGATTTCTTAGGCGCAACGAAGATCCCTTAAGCCTAGCTAGTGGGAAGGTGACT 360
Db 2613 AGGACTAGCTGGATTTCTTAGGCGCTAGCTAAGATCCCTTAAGCCTAGCTAGTGGGAAGGTGACC 2672
QY 361 GCATCCACCTCTAAACATGCGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
Db 2673 ACATCCACCTTTAAACAGCGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2732
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 480
Db 2733 ACTAAATGCTAATTAGGCAAAACAGGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 2792
QY 481 AGAGCACAGCGGAGGACCAAGGATCGGGATATAAACCAGGCAATTCGAGCCGCGCAACGG 540
Db 2793 AGAGCACAGCAGGAGGACCAATGATCGGGATATAAACCAGGCAATTCGAGCCGCGCAACGG 2852
QY 541 CAACCCCTTTGGGTCCCTCCCTTTGTATGGGGCTCTGTTTCACTCTATTTCACCTCT 600
Db 2853 CAACCCCTTTGGGTCCCTCCCTTTGTATGGGGCTCTGTTTCACTCTATTTCACCTCT 2912
QY 601 ATTAATCTTGCAACTG 617
Db 2913 ATTAATCTTGCAACTG 2929
```

RESULT 19

US-09-873-367C-81

; Sequence 81, Application US/09873367C

; Publication No. US20030165839A1

; GENERAL INFORMATION:

; APPLICANT: Young, Paul

; APPLICANT: Soppet, Daniel

; APPLICANT: Endress, Gregory

; APPLICANT: Augustus, Meena

; APPLICANT: Ebner, Reinhard

; APPLICANT: Carter, Kenneth

; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using

; FILE REFERENCE: 689290-84

; CURRENT APPLICATION NUMBER: US/09/873,367C

; CURRENT FILING DATE: 2003-04-29

; PRIOR APPLICATION NUMBER: U.S. 60/236,891

; PRIOR FILING DATE: 2000-09-29

; PRIOR APPLICATION NUMBER: U.S. 60/236,842

; PRIOR FILING DATE: 2000-09-29

; PRIOR APPLICATION NUMBER: U.S. 60/244,867

; PRIOR FILING DATE: 2000-11-01

; PRIOR APPLICATION NUMBER: U.S. 60/245,084

; NUMBER OF SEQ ID NOS: 1067

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 81

; LENGTH: 56093

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-873-367C-81

Query Match 81.3%; Score 516.2; DB 10; Length 56093;

Best Local Similarity 91.1%; Pred. No. 1.6e-149;

Matches 564; Conservative 0; Mismatches 43; Indels 12; Gaps 1;

QY 1 CCCTGTATCTTTAACTCTCTTTGTTAACTTTGTCTCTCCAGAAATCAAACTGTAATAACTA 60

Db 37274 CCCTGTATCTTTAACTCTCTTTGTTAACTTTGTCTCTCCAGAAATCGAAGCTGTAATAACTA 37333

QY 61 CAAATTGTTCTTCAAATGGAGCACCAGATGGAGTCCATGATCAAGATCCACCGTGGACCC 120

Db 37334 -----CAAATGGAGCACCAGATGGAGTCCAGACTAAGATCTACCGCAGACCC 37381

QY	121	CTGGACCGGCGCTGTAGCCCATGCTCCGATGTTAATGACATTGAAGCACCCTCCCGAG	180
Db	37382	CTGGACCGGCGCTGTAGCCCATGCTCCGATGTTAATGACATTGAAGCACCCTCCCTGAG	37441
QY	181	GAATCTCAACTGCGCAACCCCTACTATGTCGCCCAATTCAGCGGGGAAGCAGTTAGAGCGGT	240
Db	37442	GAATCTCAAGTCGCAACAACTCTACTACGCCCCCAATTCAGCAGGAGCAGTTAGAGCGGT	37501
QY	241	CATCAGCAACCTCCCCAAACAGACACTTTGGGTTTTCTGTGAGAGGGGGGACTGAGAGAC	300
Db	37502	CGTGGGCGCAACCTCCCCAAACAGACACTTAGGTTTTCTGTGAGATGGGGGACTGAGAGAC	37561
QY	301	AGGACTAGCTGGATTTCTTAGGCCAAGAAAGTCCCTAAGCCTAGCTGGGAAGGTGACT	360
Db	37562	AGGACTAGCTGGATTTCTTAGGGCTGACTAAGAAATCCCTAAGCCTAGCTGGGAAGGTGACC	37621
QY	361	GCATCCACCTCTAAACATAGGGGCTTCCAACTTAGCTCACACCCGACCAATCAGAGAGCTC	420
Db	37622	ACATCCACCTTTAAACACGGGGCTTCCAACTTAGCTCACACCTGACCAATCAGAGAGCTC	37681
QY	421	ACTAAATGCTAATTAGGCAAAAATAGGAGGTAAGAAATAGCCAAATCATCTATTGCGCTG	480
Db	37682	ACTAAATGCTAATTAGGCAAAACAGAGGAGTAAAGAAATAGCCAAATCATCTATTGCGCTG	37741
QY	481	AGACACAGCGGGAGGGAACAAGATCGGATATAAACCCAGACGATTCGAGCGCGCAACGG	540
Db	37742	AGACACAGCAGGAGGGAACAATGATCGGAGATAAAACCCAGATTCGAGCGCGCAACGG	37801
QY	541	CAACCCCTTTGGTCCCTCCCTTTGTATGGCGCTCTGTTTTCACTCATTTTCACTCT	600
Db	37802	CAACCCCTTTGGTCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCATTTTCACTCT	37861
QY	601	ATTAATCTTGCAACTGAA	619
Db	37862	ATTAATCTTGCAACTGCA	37880
RESULT 20			
US-10-632-793-30			
; Sequence 30, Application US/10632793			
; Publication No. US20040048298A1			
; GENERAL INFORMATION:			
; APPLICANT: PARANHOS-BACCALA, Glaucia			
; APPLICANT: MALLET, Francois			
; APPLICANT: VOISSET, Cecile			
; TITLE OF INVENTION: ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN			
; TITLE OF INVENTION: AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT			
; FILE REFERENCE: 110048			
; CURRENT APPLICATION NUMBER: US/10/632,793			
; CURRENT FILING DATE: 2003-08-04			
; PRIOR APPLICATION NUMBER: US/09/869,927			
; PRIOR FILING DATE: 2001-10-22			
; PRIOR APPLICATION NUMBER: PCT/FR00/00144			
; PRIOR FILING DATE: 2000-01-21			
; PRIOR APPLICATION NUMBER: FR 99/00888			
; PRIOR FILING DATE: 1999-01-21			
; NUMBER OF SEQ ID NOS: 33			
; SOFTWARE: Patentin version 3.1			
; SEQ ID NO 30			
; LENGTH: 7582			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: misc_feature			
; LOCATION: (198)..(198)			
; OTHER INFORMATION: n = a or g or c or t/u			
; FEATURE:			
; NAME/KEY: misc_feature			
; LOCATION: (307)..(307)			
; OTHER INFORMATION: n = a or g or c or t/u			
; FEATURE:			
; NAME/KEY: misc_feature			
; LOCATION: (355)..(355)			

```
Db 7504 CAACCCCTTTGGGTCCCTCCCTTTGTATGGGAGCTCTGTTTCATGCTATTTCACCTCT 7563
QY 601 ATTAAATCTTGCAACTG 617
Db 7564 ATTAAATCTTGCACTG 7580
RESULT 21
US-10-087-192-730/c
; Sequence 730, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; PRIOR FILING DATE: 2002-03-01
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 730
; LENGTH: 161334
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(161334)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-730
Query Match 80.4%; Score 510.6; DB 13; Length 161334;
Best Local Similarity 90.0%; Pred. No. 1.5e-147;
Matches 575; Conservative 0; Mismatches 49; Indels 15; Gaps 2;
QY 1 CCCTGTATCTTTAACTCTCTTTGTTAGTTTGTCTCTCCAGAAATCAAACTGTAAACTA 60
Db 100611 CCCTGTATCTTTAACTCTCTTTGTTAGTTTGTCTCTCCAGAAATCAAACTGTAA 100552
QY 61 CAAATGTTCTTCAATGGAGCAACGATGGATGCAATGATTAATGACATTCAGGCGGAGCGGT 120
Db 100551 CAAATGTTCTTCAATGGAGCAACGATGGATGCAATGATTAATGACATTCAGGCGGAG 100492
QY 121 CTGGACGGGCTCTAGCCCATGCTCCGATGTTAATGACATTCAGGCGGAGCGGT 180
Db 100491 CTGGACGGGCTCTAGCCCATGCTCCGATGTTAATGACATTCAGGCGGAGCGGT 100432
QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCCAAATTCAGGCGGAGCGGT 240
Db 100431 GAAATCTCAACTGCACAAACCCCTACTATGCCCCAAATTCAGGCGGAGCGGT 100372
QY 241 CATCAGCAACCTCCCCAACAGCATTTGGGTTTTCTGTTGAGAGGGGGAGCTGAGAGAC 300
Db 100371 CGTCAGCAACCTCCCCAACAGCATTTGGGTTTTCTGTTGAGAGGGGGAGCTGAGAGAC 100312
QY 301 AGACTAGCTGGATTTCTTAGGCAACGAGAAATCCCTTAGCCCTAGCTGGGAGAGTGACT 360
Db 100311 AGGATTAGCTGGATTTCTTAGGCGGACTAAGAAATCCCAAGCCCTAGCTGGGAGAGTGACC 100252
QY 361 GCATCCACCTCTAAACATGGGGCTTGCACCTAGCTGCACACCCGACCAATC----- 411
Db 100251 ACATCCACCTTTAAACACTGGGCTTGCACCTAGCTGCACACCCGACCAATCAGGTAGTAA 100192
QY 412 AGAGAGCTCACTAAATGCTTAATTAGGCAAAATAATAGGAGTAAAGAAATAGCAATCATC 471
Db 100191 AGAGAGCTCACTAAATGCTTAATTAGCAAAATAATAGGAGTAAAGAAATAGCAATCATC 100132
QY 472 TATTGCTGAGAGCACAGCGGGAGGACAAAGGATTCGGGATATATAAACCCAGGCAATTGAGC 531
|||||
```

```
Db 100131 TATCGCTGAGAGCACAGCGGGAGGACAAATGATCGGATATAAACCCAGGCAATTCAGC 100072
QY 532 CGCAACAGCAACCCCTTTGGGTCCCTCCCTTTGTATGGGCGCTCTGTTTTCACCTCTA 591
Db 100071 CGCAACAGGCTACCTTTCTTTGGGTCCCTCCCTTTGTATGGGAGCTCTCTCT-----GT 100018
QY 592 TTTCACTCTATTAAATCTTGCRACTGNAAAAAAAAAAAAA 630
Db 100017 CTTCACTCTATTAAATATTGCAACTGCAAAAAAAAAAATA 99979
RESULT 22
US-10-087-192-1666
; Sequence 1666, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1666
; LENGTH: 285020
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(285020)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1666
Query Match 78.9%; Score 500.8; DB 13; Length 285020;
Best Local Similarity 89.3%; Pred. No. 2.3e-144;
Matches 577; Conservative 0; Mismatches 57; Indels 12; Gaps 3;
QY 1 CCCTGTATCTTTAACTCTCTTTGTTAGTTTGTCTCTCCAGAAATCAAACTGTAAACTA 60
Db 279486 CCCTGTATCTTTAACTCTCTTTGTTAGTTTGTCTCTCCAGAAATCGAAGCTGTAAACTA 279545
QY 61 CAAATGTTCTTCAATGGAGCAACGATGGAGTCCATGATTAATGACATTCAGGCGGAGCGGT 120
Db 279546 CAAATGTTCTTCAATGGAGCAACGATGGAGTCCATGATTAATGACATTCAGGCGGAGCGGT 279605
QY 121 CTGGACGGGCTCTAGCCCATGCTCCGATGTTAATGACATTCAGGCGGAGCGGT 180
Db 279606 CTGGACGGGCTCTAGCCCATGCTCCGATGTTAATGACATTCAGGCGGAGCGGT 279665
QY 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCCAAATTCAGGCGGAGCGGT 240
Db 279666 GAACTCTCAACTGCACGACCCCTACTATGCCCCAAATTCAGGCGGAGCGGT 279725
QY 241 CATCAGCAACCTCCCCAACAGCATTTGGGTTTTCTGTTGAGAGGGGGAGCTGAGAGAC 300
Db 279726 CATTCGCAACCTCCCCAACAGCATTTGGGTTTTCTGTTGAGAGGGGGAGCTGAGAGAC 279785
QY 301 AGGACTAGCTGGATTTCTTAGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 358
Db 279786 AGGACTAGCTGGATTTCTTAGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 279845
QY 359 CTGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCAACCCGACCAATC----- 411
Db 279846 CGGCATCCACCTTTAAACAGCGGGCTTGCAACTTAGCTCAACCCGACCAATCAGGAGT 279905
QY 412 --AGAGAGCTCACTAAATGCTTAATTAGGCAAAATAATAGGAGTAAAGAAATAGCCCAATCA 469
|||||
```

Db 279906 AAGACAGCTCACTAAATGCTAATAGGCTAAACAGGAGTAAAGAAATAGCCAATCA 279965
Qy 470 TCTATTGCTGAGAGACAGCGGAGGACAGGATCGGGATATAAACCCAGGCAATTCGA 529
Db 279966 TCTATCGCTGAGAGCAC -GAGGAGGACAAATGATCAGGATATAAACCCAGGCAATTCAA 280024
Qy 530 GCCGGCAACGGCAACCCCTTTGGTCCCTCTCTTTGATGGCGCTCTGTTTTCACATC 589
Db 280025 GCCGGCAGTGGCTACCGCTTTGGGTCCCTCTGTTTGTATGGAGCTCTGTTTTCACATC 280084
Qy 590 TATTTCACCTCTATTAATCTGCAACTGCAAAAAAAGAAAAAAGAAAAA 635
Db 280085 TATTAAATCTTGCAACTGCAAAAAAAGAAAAAAGAAAAAAGAAAAA 280130

RESULT 23
US-09-864-761-4444
; Sequence 4444, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aesomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864, 761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4444
; LENGTH: 1894
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002346.1

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
US-09-864-761-4444
Query Match 78.7%; Score 499.6; DB 9; Length 1894;
Best Local Similarity 88.7%; Pred. No. 5.2e-145;
Matches 571; Conservative 0; Mismatches 54; Indels 19; Gaps 2;
Qy 1 CCTGTATCTTTAACTCTCTTTGTTAAAGTTCTCTCTCCAGAATCAAAATCTGTAATACTA 60
Db 1217 CCTGTATCTTTAACTCTCTTTGTTAAAGTTCTCTCTCCAGAATCGAAGTAAACTA 1276
Qy 61 CAAATTTGTTTCAAAATGGAGCACAGATGGAGTCCATGACTAAGATCCACGTTGGACCC 120
Db 1277 CAAATCGTTCTTCAAAATGGAGCCCGCCAGATGCATGCATGAGTAAATCTACCCAGGCC 1336
Qy 121 CTGGACCGGCTCTGAGCCCATCTCCGATGTTTAAATGATTAAGATGAGGACCCCTCCCGAG 180
Db 1337 CTGGACCGGCTCTGAGCCCATCTCTGATGTTAATGATCAATCAAAAGGACCCCTCCCGAG 1396
Qy 181 GAAATCTCAACTGCAACACCCCTCTATGCCCCCAATTCAGCGGGAGGAGCTAGAGCGGT 240
Db 1397 GAAATCTCAACTGCAACACCTCTATAGCCCCCAATTCAGCGGGAGGAGCTAGAGTGGT 1456
Qy 241 CATCAGCCAACTCCCCAACAGCACCTTGGGTTTTCTGTTGAGAGGGGGAGCTGAGAGAC 300
Db 1457 TGTGGCCAACTCCCCAACAGCACGTTGGGTTTTCTGTTGAGAGGGGGAGCTGAGAGAC 1516
Qy 301 AGACTAGCTGATTTCTTAGGCGCAACGAGAAATCCCTAAGCTAGCTGGGAAAGTGAAT 360
Db 1517 AGGAAATACTAGATTTCTTAGACCAACTTGGGTTTTCTGTTGAGAGGGGGAGCTGAGAGAC 1576
Qy 361 GCATCCACCTCTAAACATGGGGCTTGCACACTTAGCTCACACCCGACCAATC----- 411
Db 1577 GCTTCACCTTTAAACACCGCGCTTGCACACTTAGCTCACACCCGACCAATCACTAA 1636
Qy 412 AGAGAGCTCACTAAATGCTAATTAGGCAAAATAGGAGGTAAGAAATAGCAATCATC 471
Db 1637 AGAGAGCTCACTAAATGCTAATTAGGCAAAATAGGAGGTAAGAAATAGCAATCATC 1696
Qy 472 TATTGCTGAGAGCACAGCGGGAGGACAAAGATCGGATATATAAACCAGGCAATTCGAGC 531
Db 1697 TGTGCTGACAGCACAGGAGGAGCAATGATCGGATATATAAACCAGGCAATTCGAGC 1756
Qy 532 CGGCAACGGCAACCCCTTTGGGTCCCTCTCCCTTTGATGGGCTCTGTTTTCATCTA 591
Db 1757 CAGCTACAGCTACCTCTTTGGGTCCCTCTCCCTTTGATGGGAGCTCTGT----- 1806
Qy 592 TTTCACTCTATTAAATCTTGCAACTGAAAAAAGAAAAAAGAAAAA 635
Db 1807 CTTCACTCTATTAAATCTTGCAACTGAAAAAAGAAAAAAGAAAAA 1850

RESULT 24
US-10-632-793-28
; Sequence 28, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: MALLET, Francois
; APPLICANT: VOISSET, Cecile
; TITLE OF INVENTION: ENDOGENOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
; TITLE OF INVENTION: AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
; FILE REFERENCE: 110048
; CURRENT APPLICATION NUMBER: US/10/632,793

```

; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: US/09/869,927
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: PCT/FR00/00144
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: FR 99/00888
; PRIOR FILING DATE: 1999-01-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 28
; LENGTH: 3372
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-632-793-28

Query Match      78.0%; Score 495.6; DB 16; Length 3372;
Best Local Similarity 92.2%; Pred. No. 1.2e-143;
Matches 522; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 70 CTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCCCTGGGACCGG 129
DB 2802 CTAATAATGGAGCCCAAGATGCAATCCAGACTAAGATCTACCGCAGACCCCTGGACCGG 2861

QY 130 CTTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCAACCCCTCCCGAGGAAATCTCA 189
DB 2862 CTTGTTAGCCACGATCTGATGTTAATGACATCAAGGCAACCCCTCTGAGGAAATCTCA 2921

QY 190 ACTGCACAACCCCTACTATGCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCTATCAGCCA 249
DB 2922 GCTGCACAACCTCTACTACGCCCAATTCAGCAGGAAGCAGTTAGAGCGGTCTCGGCCA 2981

QY 250 ACCTCCCAACAGCAGCTTTGGTGTTCCTGTGAGAGGGGGAAGCTGAGAGACAGGACTAGC 309
DB 2982 ACCTCCCAACAGCAGCTTAGGTTTCTGTGAGATGGGGGAAGCTGAGAGACAGGACTAGC 3041

QY 310 TGGATTTCTAGCCACGAGGATCCCTAAGCCTAGCTGGGAGGCTGATGATCCACC 369
DB 3042 TGGATTTCTAGGCTGATTAAGGATCCCTAAGCCTAGCTGGGAGGCTGATGATCCACC 3101

QY 370 TCTAAACATGGGCTTGCACTTAGCTCACACCCGACCAATCAGAGAGCTCACTAAATG 429
DB 3102 TTTAAACAGGGCTTGCACTTAGCTCACACCTGACCAATCAGAGGCTCACTAAATG 3161

QY 430 CTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTTCCTGAGAGCAGACAG 489
DB 3162 CTAATTAGGCAAAAGCAGGAGGTAAAGAAATAGCAATCATCTATTTCCTGAGAGCAGACAG 3221

QY 490 CGGAGGGCAAGGATCGGATATAACCCAGGCAATTCGAGCCGCAACCGCAACCCCT 549
DB 3222 CAGGAGGGCAATGATCGGATATAACCCAGGTTTTCGAGCCGCAACCGCAACCCCT 3281

QY 550 TTGGGTCCCTCCCTTTGTATGGCGCTCTGTGTTTCACTCTATTTCACCTCTATTAAATCT 609
DB 3282 TTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTTCATGCTATTTCACCTCTATTAAATCT 3341

QY 610 TGCACACTGAAAAAATAAAAAAAAAAAAAA 635
DB 3342 TGCACACTGCAAAAAAATAAAAAAAAAAAAAA 3367

RESULT 25
US-10-276-774-678
; Sequence 678, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y, Tom et al
; TITLE OF INVENTION: No. US20040053245A1e1 Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27

```

```

; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 678
; LENGTH: 2052
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-276-774-678

Query Match      77.7%; Score 493.2; DB 16; Length 2052;
Best Local Similarity 91.2%; Pred. No. 5.3e-143;
Matches 549; Conservative 0; Mismatches 43; Indels 10; Gaps 2;

QY 1 CCTGTATCTTTAAACCTCTTGTAAAGTTTGTCTTCCAGAAATCAAAACTGTAATACTA 60
DB 993 CCTGTATCTTTAAACCTCTTGTAAAGTTTGTCTTCCAGAAATCAAAAGCTGTAATACTA 1052

QY 61 CAAATTGTTCTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
DB 1053 CACATCGTTCTTCAATGGAGCCCGCAGATGCAGTCCATGACTAAGATCTACCGCGATCC 1112

QY 121 CTGGACCGGCTGTAGCCCATGCTCCGATGTTAATGACATTGAAGGCAACCCCTCCCGAG 180
DB 1113 CTGGACCGGCTGTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACTCTCTCCGAG 1172

QY 181 GAAATCTCAACTGCAACACCCCTACTATGCCCAATTCAGCGGGAAGCAGTTAGACGGT 240
DB 1173 GAAATCTCAATGCAACACCCCTACTATGCCCAATTCAGCAGGAAGCAGTTGGAGCAGT 1232

QY 241 CATCAGCAACCTCCCAACAGCAGCTTGGGTTTCTGTGAGAGGGGGGAGCTGAGAGAC 300
DB 1233 CGACGGCAACCTCCCAACAGCAGCTTGGGTTTCTGTGAGAGGGGGTACTGAGAGAC 1292

QY 301 AGGACTAGCTGGATTTCTTAGGCCAACGAAAGATCCCTAAGCCTAGCTGGGAAGGTGACT 360
DB 1293 AGGACTAGCTGGATTTCTTAGGCCGATGAAGATCCCTAAGCCTAGCTGGGAAGGTGACT 1352

QY 361 GCATCCACCTTAAACATGGGCTTGCAACTTAGCTCACACCCGACCAATC----- 411
DB 1353 GTATCCACCTTTAAACACAGGGCTTGCAACTTAGCTCACACCCGACCAATCAGGTAGTAA 1412

QY 412 AGAGAGCTCACTAAATGCTAATTAGGCAAAATAGGAGGTAAGAAATAGCAATCATC 471
DB 1413 AGAGAGCTCACTAAATGCTAATTAGGCAAAATAGGAGGTAAGAAATAGCTAAT-ATC 1471

QY 472 TATTGCTCTAGAGCAGCAGCGGAGGACAAAGGATCGGATATATAACCCAGGCAATTCGAGC 531
DB 1472 TATCAGCTGAGAGTACAGGGGAGGAGCAATGATTTGGGATAGAAACCCAGGCAATTCGAGC 1531

QY 532 CGGCAACGGCAACCCCTTTGGGTCCCTCCCTTTGTATGGCGCTCTGTGTTTCACTCTA 591
DB 1532 CGGCAACGGCAACCCCTTTGGGTCTCTCTTCCATTTTATGGGAGCTCTGTGTTTCACTCTA 1591

QY 592 TT 593
DB 1592 TT 1593.

RESULT 26
US-09-997-722-148
; Sequence 148, Application US/09997722
; Publication No. US20040072154A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71171/RMS/DCF
; CURRENT APPLICATION NUMBER: US/09/997,722
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586

```

```

; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 148
; LENGTH: 22436
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-997-722-148

Query Match      76.0%; Score 482.6; DB 11; Length 22436;
Best Local Similarity 87.6%; Pred. No. 3.4e-139; Indels 11; Gaps 3;
Matches 565; Conservative 0; Mismatches 69;

QY      1  CCTGTATCTTTAAACCTCTTCTTAAAGTTTGTCTTCCAGAATCAAACTGTAAACTA 60
DB      5655 CCTGTACTTTAAACCTCTTCTTAAAGTTTGTCTTCCAGAATCGAAGCTGTAAACTA 5714

QY      61  CAAATGTTCTTCAATGAGACACGATGGAGTCATGACTAGATCCACCGTGGACCC 120
DB      5715 CAAATGTTCTTCAATGAGACCCGATGCGATGCACTGATGACTAGATCTATCGAGACCC 5774

QY      121 CTGACCGGCTCTAGCC- CATGCTCCGATGTTAATGACATTTGAAGCACCCCTCCCGA 179
DB      5775 CTGACTGGCTCTAGCTCATGCTCCCATGTTAATGACATTTGAAGCAACTCTCCCGA 5834

QY      180 GGAATCTCAACTGCAACCCCTACTATGCCCAATTGAGCGGAACGATGAGCGG 239
DB      5835 GGAATCTCAACTGCAACCCCTACTACACCCCAAGTTGAGCGGAACGATGAGCA 5894

QY      240 TCATCAGCAACCTCCCAACAGCACTTGGTTTTCCTGTCAGAGGGGGGACTGAGAGA 299
DB      5895 TCATTGGCCCACTCCCAACAGCACTTGGTTTTCCTGTCAGAGGGGGGACTGAGAG- 5953

QY      300 CAGGACTAGCTGGATTTCTTAGGCCAACGAGAAATCCCTTAAGCTAGCTGGAGGTGAC 359
DB      5954 CAGGACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCCTTAAGCTAGCTGGAGGTGAC 6013

QY      360 TGCATCAGCTTAAACAGCTGGGCTTGCACTTAGCTCACACCGACCAATC----- 411
DB      6014 TGCACCCACCTTTTAAACAGGCTTGCACTTAGCTCACCTGACCAATCAAGTAGTA 6073

QY      412 -AGAGAGCTCACTAAATGCTAATAGGCAAAATAGGAGGTAAAGAAATAGCAATCAT 470
DB      6074 AAGAGAGCTCACTAAATGCTAATAGGCAAAATAGGAGGTAAAGAAATAGCAATCAT 6133

QY      471 CTATTGCTGAGACAGCGGGAGGACAGATCGGATATATAAACCAGGCAATTCGAG 530
DB      6134 CTATCAGCTGAGACAGCGGTGGGAGCAATGATCAGATATATAAACCAGGCAATTCGAG 6193

QY      531 CCGCAACGGCAACCCCTTTGGGTCCTCCCTTGTATGGCGCTCTGTTTCACTCT 590
DB      6194 CCAGCAATGGCTACCTCTTTGGGTCCTCCCTTGTATGGGAGCTCTGTTTACACTCT 6253

QY      591 ATTCACTCTATTAAATCTGCACTGAAATGAAAAA 635
DB      6254 ATTAAACCTTCAACTGCAAAATATATATATATATATATGAAATA 6298

RESULT 27
US-10-220-120-15
; Sequence 15, Application US/10220120
; Publication No. US2004048253A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: PANZER, Scott R.
; APPLICANT: SPIRO, Peter A.
; APPLICANT: BANVILLE, Steven C.
; APPLICANT: SHAH, Purvi
; APPLICANT: CHALUP, Michael S.
; APPLICANT: CHANG, Simon C.
; APPLICANT: CHEN, Alice
; APPLICANT: D'SA, Steven A.
; APPLICANT: AMSHEY, Stefan

; APPLICANT: DAHL, Christopher R.
; APPLICANT: DAM, Tam C.
; APPLICANT: DANIELS, Susan B.
; APPLICANT: DUFOUR, Gerard E.
; APPLICANT: FLORES, Vincent
; APPLICANT: FONG, Willy T.
; APPLICANT: GREENAWALT, Lila B.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: JONES, Anissa L.
; APPLICANT: LIU, Tommy F.
; APPLICANT: ROSEBERRY, Ann M.
; APPLICANT: ROSEN, Bruce H.
; APPLICANT: RUSSO, Frank D.
; APPLICANT: STOCKDREHER, Theresa K.
; APPLICANT: DAFRO, Abel
; APPLICANT: WRIGHT, Rachel J.
; APPLICANT: YAP, Pierre E.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: BRADLEY, Diana L.
; APPLICANT: BRATCHER, Shawn R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: COHEN, Howard J.
; APPLICANT: HODGSON, David M.
; APPLICANT: LINCOLN, Stephen E.
; APPLICANT: JACKSON, Stuart
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PT-1113 PCT
; CURRENT APPLICATION NUMBER: US/10/220,120
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 60/184,777; 60/184,797; 60/184,698; 60/184,770; 60/184,774;
; 60/184,693; 60/184,771; 60/184,813; 60/184,773; 60/184,776;
; 60/184,769; 60/184,768; 60/184,837; 60/184,597; 60/184,841;
; 60/184,772; 60/185,213; 60/185,216; 60/204,863; 60/205,221;
; 60/204,815; 60/203,785; 60/204,821; 60/204,908; 60/204,226;
; 60/204,525; 60/205,285; 60/205,232; 60/205,287;
; 60/205,324; 60/205,286
; PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
; 2000-05-17; 2000-05-12; 2000-05-16; 2000-05-15;
; 2000-05-17; 2000-05-17; 2000-05-17; 2000-05-17; 2000-05-17;
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: PERL Program
; SEQ ID NO 15
; LENGTH: 849
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAMES/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US2004048253A1 LG:429446.1:2000FEB18
US-10-220-120-15

Query Match      75.8%; Score 481.4; DB 16; Length 849;
Best Local Similarity 89.1%; Pred. No. 1.7e-139; Indels 9; Gaps 1;
Matches 534; Conservative 0; Mismatches 56;

QY      27  GTTTGTCTCTTCCAGATCAAAACTGTAAACTACAAATTTGTTCTTCAATGGAGCA 86
DB      121 GTTTGTCTCTTCCAGATCAAAACTGTAAACTACAAATTTGTTCTTCAATGGAGCA 180

QY      87  GATGAGTCCATGATTAAGATCCACCGTGGACCCCTGGACCGCTGTAGCCCATGCTC 146
DB      181 GATGAGTCTATGACTAAGATCTTACCACAGACCTTTGGACCGGCTGTAGCCCATGCTC 240

QY      147  CGATGTTAATGACATTCGAAGCGACCCCTCCCGAGGAAATCTCAATGCAACACCCCTACT 206
DB      241 CGATGTTGATGACATCGAAGCGCCCTCCCAAGGAAATCTCAATGCAATGACCCCTACT 300

QY      207  ATGCCCAATTCAGCGGGAAGAGCTTAGCGGTATAGCGGTATCAGCCAACTCCCAACAGCACT 266
DB      301 ACGCCCAATTCAGCGGGAAGAGCTTAGAGAGGTATAGAGAGGTATTCAGCCAACTCCCAACAGCACT 360

```

QY 267 TGGGTTTCTGTTGAGAGGGGGAGCTGAGAGACAGGACTAGCTGGATTTCTTAGGCCAA 326
DB 361 TGAAGTTTCTGTTGAGAGGGGAACTGAGAGACAGGACTAGCTGGATTTCTTAGGCCAA 420
QY 327 CGAAGAAATCCTTAAGCCTAGCTGGGAAGGTGACTGCATCCACCTCTAAACATGGGGCTTG 386
DB 421 CTAAAGAAATCCTTAAGCCTATCTGGGAAGGTGACCGTATCCACTTTAAACATAGGGCTTG 480
QY 387 CAATCTAGCTCACACCCGACCAATCA-----GAGAGCTCACTAAAAATGCTAATTAG 437
DB 481 CAATCTAGCTCACACCCGACTAATCAGATAGTAAGGAGAGCTCACTAAAAATGCTAATTAG 540
QY 438 GCAAAATAGGAGGTAAAGAAATAGCCAAATCATCTATTGCTGAGAGCACAGGGAGGG 497
DB 541 GCAAAATAGGAGGTAAAGAAATAGCCAAATCATCTATTGCTGAGAGCACAGGAGGG 600
QY 498 ACAAGATCGGATATAAACCCAGGCAATTCGAGCGGCAACGCCCTTTGGGTCC 557
DB 601 ACAATGATCAGGATATAAACCCAGGCAATTCGAGCGCAATGCTACCTCTTTGGGTCC 660
QY 558 CCTCCCTTTGATGGGGCTCTGTTTTCACCTCTATTTCACCTATTAAATCTTGCAACT 616
DB 661 CCTCCGTTTGTATGGGAGCTCTGTTTTCATGCTATTTTAAAGTCTTGCAACTGCACACT 719
RESULT 28
US-10-104-047-1148
; Sequence 1148, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1el full length cdna
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1148
; LENGTH: 2349
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1148
Query Match 73.4%; Score 466; DB 15; Length 2349;
Best Local Similarity 86.4%; Pred. No. 1.8e-134;
Matches 541; Conservative 0; Mismatches 75; Indels 10; Gaps 2;
QY 19 CTTGTTAAGTTTGTCTCTCCAGAATCAAACTGTAATACTAATAATGTTTCTCAAATG 78
DB 1377 CTTCTTCAGTTTGTCTCTCCAGAATCAAACTGTAATACTAATAATGTTTCTCAAATG 1436
QY 79 GAGACACAGATGAGTCCATGACTAAGATCCAGTGGACCCCTGACCGGCTCTGTAGC 138
DB 1437 GAGCCACAGATGAGTCCATGACTAAGATCCAGTGGACCCCTGACCGGCTCTGTAGC 1496
QY 139 CCATGCTCCGATGTTAATGACTTAAGGACCCCTCCCGAGGAATCTCAATGTCACAA 198
DB 1497 CCA-GCTCCACATTAATGACTTAAGGACCCCTCCCGAGGAATCTCAATGTCATGA 1555
QY 199 CCCTACTATGCCCAATTCAGCGGGAAGCAGTTAGAGCGGTGATCAGCCAACTCCCCA 258
DB 1556 CACTACTAGCCCCAATTCAGCAGGAAGCAGTTAGTGTGTTGGCCAACTCCCA 1615
QY 259 ACAGCACTTGGTTTCTGTTGAGAGGGGAGTCTGAGACAGGACTAGCTGGAATTC 318
DB 1616 ACAGCACTTGGTTTCTGTTGAGAGGGGAGTCTGAGACAGGACTAGCTGGAATTC 1675
QY 319 TAGGCCCAACCAAGATCCCTAAGCTAGTGGGAAGTGTGCTGCATCCACCTTAAACAT 378
DB 1676 TAGGCCCAACCAAGATCCCTAAGCTAGTGGGAAGTGTGCTGCATCCACCTTAAACAT 1735

QY 379 GGGGCTTGCAACTTAGCTCACACCCGACCAATC-----AGAGAGCTCACTAAATG 429
DB 1736 TAGGCTTGCAACTTAGCTCACACCCGACCAATCAGGTAGTAAAGAGAGCTTGTCTAAATG 1795
QY 430 CTAAATAGGCAAAATAGGAGGTAAAGAAATAGCCAAATCATCTATTGCTGAGAGCACAG 489
DB 1796 CTAAATAGGCAAAATAGGAGGTAAAGAAATAGCCAAATCATCTATTGCTGAGAGCACAA 1855
QY 490 CGGAGGGGCAAGGATCGGGATATAAACCCAGGCAATTCGAGCGGCAACGCCCT 549
DB 1856 GGGGCGGCAATGATCAGGATATAAACTCAGGCAATTCAGCCAGCAATGGCTACCCACT 1915
QY 550 TTGGGTCCCTCCCTTTGTTATGGGCGCTCTGTTTTCACCTCTATTTCATCTAATCT 609
DB 1916 TTGGGTCCCTCCCTTTTATGGGAGCTCTGTTTTCACCTCTATTTCATCTAATCT 1975
QY 610 TGCAACTTGAAATTAATAAATAAATAAATAA 635
DB 1976 AAAAAAAAAAAAAAAAAAGAAAAA 2001
RESULT 29
US-10-204-887-10
; Sequence 10, Application US/10204887
; Publication No. US20030124569A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: PANZER, Scott R.
; APPLICANT: SPIRO, Peter A.
; APPLICANT: BANVILLE, Steven C.
; APPLICANT: SHAH, Purvi
; APPLICANT: CHALUP, Michael S.
; APPLICANT: CHEN, Simon C.
; APPLICANT: CHEN, Alice
; APPLICANT: D'SA, Steven A.
; APPLICANT: AMSHEY, Stefan
; APPLICANT: DAHL, Christopher R.
; APPLICANT: DAM, Tam C.
; APPLICANT: DANIELS, Susan E.
; APPLICANT: DUFOUR, Gerard E.
; APPLICANT: FLORES, Vincent
; APPLICANT: FONG, Willy T.
; APPLICANT: GREENAWALT, Lila B.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: JONES, Anissa L.
; APPLICANT: LIU, Tommy F.
; APPLICANT: ROSEBERRY, Ann M.
; APPLICANT: RUSSO, Frank D.
; APPLICANT: STOCKDREHER, Theresa K.
; APPLICANT: DAFFO, Abel
; APPLICANT: WRIGHT, Rachel J.
; APPLICANT: YAP, Pierre E.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: BRADLEY, Diana L.
; APPLICANT: BRATCHER, Shawn R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: COHEN, Howard J.
; APPLICANT: HODGSON, David M.
; APPLICANT: LINCOLN, Stephen E.
; TITLE OF INVENTION: SECRETORY MOLECULES
; FILE REFERENCE: PT-1134 PCT
; CURRENT APPLICATION NUMBER: US/10/204,887
; PRIOR FILING DATE: 2002-08-21
; PRIOR APPLICATION NUMBER: 60/185,216; 60/185,215; 60/185,216; 60/205,323; 60/205,287;
; 60/205,324; 60/205,286
; PRIOR FILING DATE: 2000-02-24; 2000-05-16; 2000-05-17; 2000-05-17;
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PERL Program
; SEQ ID NO 10
; LENGTH: 1564

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030124569A1 LI:813218.1:2000FEB01
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 307, 317-318, 335, 918, 921, 943, 1011
; OTHER INFORMATION: a, t, c, g, or other
US-10-204-887-10

Query Match 70.3%; Score 449; DB 15; Length 1564;
Best Local Similarity 90.3%; Pred. No. 3e-129;
Matches 490; Conservative 0; Mismatches 45; Indels 4; Gaps 1;

Qy 41 GAATCAAACTGTAAACTACAAATGTTCTTCAATGGAGACACAGATGGAGTCCATGA 100
Db 1022 GAATTGAACTGTAAACTACAAATGTTCTTCAATGGAGACACAGATGGAGTCCATGA 1081

Qy 101 CTAAGATCCACCGTGGACCCCTGGACCGGCTGCTAGCCCATGCTCCGATGTTAATGACA 160
Db 1082 CTAAGATCCACCGTGGACCCCGGACCGGCTCTCCAGCCCATGCTCTGGTGTAAATGACA 1141

Qy 161 TTGAAGGCACCCCTCCCGAGGAAATCTCAACTGTCACAAACCCCTACTATGCCCCCAATTCAG 220
Db 1142 TCGNAGGCACCCCTCCCAAGGAAATCTCAGCTGCACAAACCCCTCTATGCCCCCAGTTCAG 1201

Qy 221 CGGGAAGCAGTTAGACGGTCTATCAGCCAACTCCCAACAGACATTTGGTCTTCTGTT 280
Db 1202 CAGGAAGCAGTTAGACGAGTCTATCGGCCAACTCCCAATAGCATTGGGTTTTCTCTGTT 1261

Qy 281 GAGGGGGGACTCAGACAGGAGTACTAGTGGATTTCTTAGGCCAACAGAAATCCCTAA 340
Db 1262 GAGAGTGGGGACTCAG----AGGACTAGCTGGATTTCTTAGGCCACTAGAAATCCCTAA 1317

Qy 341 GCCTAGCTGGGAAGTGAAGTGCATCCACCTCTAAACATGGGGTGTGCAATTAGCTCACA 400
Db 1318 GCCTAGCTGGGAAGTGAAGTGCATCCATCTTTAAACATGGGGTGTGCAATTAGCTCACA 1377

Qy 401 CCCGACCAATCAGAGAGTCTCTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAAT 460
Db 1378 CCCAACCAATCAGAGAGTCTCTAAATGCTAATTAGGCAAAACAGGAGGTAAAGAAAT 1437

Qy 461 AGCCAACTCATCTATTGCTGAGACACAGCGGGAGGACAAAGATCGGGATATAAACCACA 520
Db 1438 AGCCAACTCATCTATTGCTGAGACACAGCGGGAGGACAAAGATCAGGATATATAATCCA 1497

Qy 521 GGCATTTCAGCGGCAACGGCAACCCCTTTTGGGTCCCTCCCTTTGTATGGGCGCTCT 579
Db 1498 GGCATTTCAGCGCAGCAATGGCAACCCCTTTTGGGTCCGCTCTCTGTATGGGAGCTCTAT 1556

RESULT 30
US-10-220-120-17
; Sequence 17, Application US/10220120
; Publication No. US20040048253A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: PANZER, Scott R.
; APPLICANT: SPIRO, Peter A.
; APPLICANT: BANVILLE, Steven C.
; APPLICANT: SHAH, Purvi
; APPLICANT: CHALUP, Michael S.
; APPLICANT: CHANG, Simon C.
; APPLICANT: CHEN, Alice
; APPLICANT: D'SA, Steven A.
; APPLICANT: AMSHEY, Stefan
; APPLICANT: DAHL, Christopher R.
; APPLICANT: DAM, Tam C.
; APPLICANT: DANIELS, Susan E.
; APPLICANT: DUFOUR, Gerard E.
; APPLICANT: FLORES, Vincent
; APPLICANT: FONG, Willy T.
```

```
; APPLICANT: GREENAWALT, Lila B.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: JONES, Anissa L.
; APPLICANT: LIU, Tommy F.
; APPLICANT: ROSEBERRY, Ann M.
; APPLICANT: ROSEN, Bruce H.
; APPLICANT: RUSSO, Frank D.
; APPLICANT: STOCKDREHER, Theresa K.
; APPLICANT: DAFFO, Abel
; APPLICANT: WRIGHT, Rachel J.
; APPLICANT: YAP, Pierre E.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: BRADLEY, Diana L.
; APPLICANT: BRATCHER, Shawn R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: COHEN, Howard J.
; APPLICANT: HODGSON, David M.
; APPLICANT: LINCOLN, Stephen E.
; APPLICANT: JACKSON, Stuart
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PT-1113 PCT
; CURRENT APPLICATION NUMBER: US/10/220,120
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 60/184,777; 60/184,797; 60/184,698; 60/184,770; 60/184,774;
60/184,693; 60/184,771; 60/184,813; 60/184,773; 60/184,776;
60/184,769; 60/184,768; 60/184,837; 60/184,897; 60/184,841;
60/184,772; 60/185,213; 60/185,216; 60/204,863; 60/205,221;
60/204,815; 60/203,785; 60/204,821; 60/204,908; 60/204,226;
60/204,525; 60/205,285; 60/205,232; 60/205,323; 60/205,287;
60/205,324; 60/205,286
; PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
2000-05-17; 2000-05-12; 2000-05-16; 2000-05-16; 2000-05-15;
2000-05-16; 2000-05-17; 2000-05-16; 2000-05-17; 2000-05-17;
2000-05-17; 2000-05-17
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: PERL Program
; SEQ ID NO 17
; LENGTH: 1393
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040048253A1 LI:351965.1:2000FEB01
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 120, 765
; OTHER INFORMATION: a, t, c, g, or other
US-10-220-120-17
```

```
Query Match 70.3%; Score 446.2; DB 16; Length 1393;
Best Local Similarity 90.3%; Pred. No. 2.2e-128;
Matches 502; Conservative 0; Mismatches 44; Indels 10; Gaps 2;

Qy 47 AAACGTGTAACCTACAAATTTGTTCTTCAATGAGACACAGATGGAGTCCATGACTAAGA 106
Db 2 AAGCTGTAAACTACTAAATTTGTTCTTCAAAAGGAGGCCAGATGCAGTCCATGACTAAGA 61

Qy 107 TCCACCGTGGACCCCTGGACCGGCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAG 166
Db 62 TCTACACACAGACCCCTGGACCGGCTGCTAGCCCATGCTAGCCGATGTTAATGATATGANA 121

Qy 167 GCACCCCTCCCGAGGAAATCTCAACTGACAAACCCCTACTATATGCCCAATTCAGCGGAA 226
Db 122 GCATCC-CCCAAGGAAATTTCAACTGACAAACCCCTACTACACCCCAATTCAGCAGAA 180

Qy 227 GCAGTTAGAGGGTTCATCAGCCCAACCTCCCAACAGCAGCTGGGTTTTCTGTTGAGAG 286
Db 181 GCAGTTAGAGGGTTCGTGAGCCCAACCTCCCAACAGCAGCTGGGTTTTCTGTTGAGAGC 240

Qy 287 GGGGACTGAGAGACAGGACTAGCTGGATTTCTTAGGCCAACGACGAATCCCTAAGCCTAG 346
```

```
|||||
Db 241 GGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGATCCCTAAGCCTAG 300
Qy 347 CTGGGAAGGTGACTGCATCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGAC 406
Db 301 CTGGGAAGGTGACCGGCTCTACCTTTAAACACACGGGGCTTGCAACTTAGCTCACACCCAAC 360
Qy 407 CAATC-----AGAGAGCTCACTAAATGCTTAATTAGGCAAAATAGGAGGTAAGA 457
Db 361 CAATCAGGTAGTAAAGAGAGCTCACTAAATGCTTAATTAGGCAAAACACGGAGGTAGAGA 420
Qy 458 AATAGCCAATCATCTATTGCTGAGAGCACAGCGGGAGGACAAAGGATCGGGATATAAAC 517
Db 421 AATAGCCAATCACTATCGCTGAGAGCACAGCGGGAGGACAAATGATCCGGATATAAAC 480
Qy 518 CCAAGCATTTCGAGCCGCAACGGCAACCCCTTTGGGTCCCTTCCCTTTGTATGGGGGCT 577
Db 481 CCAAGCATTTCGAGCCGCAACGGCTGCCCTTTGTGTCCCTCCCTTTGTATGGGAGCT 540
Qy 578 CTGTTTCACTCTATT 593
Db 541 CTGTTTCACTCTATT 556
```

Search completed: January 22, 2005, 20:08:26
Job time : 355.872 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2005, 12:01:19 ; Search time 1891.54 Seconds
(without alignments)
12232.983 Million cell updates/sec

Title: US-09-319-156b-6
Perfect score: 635
Sequence: 1 cctgtatcttttaacctctt.....tgaaaaaaaccccccccccc 635

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	498.8	78.6	1160	3 BC030968	BC030968 Homo sapi
2	492.8	77.6	2748	3 CR605851	CR605851 full-leng
3	488.8	77.0	2749	3 CR617248	CR617248 full-leng
4	488	76.9	2500	3 CR622175	CR622175 full-leng
5	485.6	76.5	771	4 BI087886	BI087886 602852690
6	485	76.4	2748	3 CR625046	CR625046 full-leng
7	470.6	74.1	719	7 CN272394	CN272394 170006000
8	468.6	73.8	1071	5 BX365066	BX365066 BX365066
9	466.6	73.5	758	5 BX357208	BX357208 BX357208
10	464.6	73.2	2716	3 CR613169	CR613169 full-leng
11	454.2	71.5	998	5 BX337769	BX337769 BX337769
12	453.6	71.4	1058	5 BX378303	BX378303 BX378303
13	450.8	71.0	689	9 AG121669	AG121669 Pan trogl
14	437	68.8	494	1 AA781423	AA781423 a126c03.s
15	435	68.5	1019	5 BX439636	BX439636 BX439636
16	432.4	68.1	522	2 AW971553	AW971553 EST383642
17	430.2	67.7	870	7 CN645411	CN645411 ILLUMIN
18	424	66.8	1500	3 BC026287	BC026287 Homo sapi
19	423	66.6	653	9 AG033781	AG033781 Pan trogl
20	415.8	65.5	490	1 AF598135	AF598135 t014a10.x
21	397.4	62.5	440	2 BE732673	BE732673 601571305
22	396	62.4	543	4 BI933185	BI933185 ie57c10.x
23	395.2	62.2	966	5 BX380176	BX380176 BX380176
24	394.8	62.2	651	9 AG058970	AG058970 Pan trogl

C 25	393.2	61.9	701	9	AG126669	AG126669 Pan trogl
C 26	393	61.9	609	9	AG066901	AG066901 Pan trogl
C 27	392.6	61.8	443	1	AA837267	AA837267 od26b10.s
C 28	387.2	61.0	712	8	AQ892947	AQ892947 HS_3131_B
C 29	386.8	60.9	436	1	AI128526	AI128526 qc61h10.x
C 30	385.2	60.7	446	1	AI393478	AI393478 tg45g04.x
C 31	384.6	60.6	641	9	AG036829	AG036829 Pan trogl
C 32	384.2	60.5	485	2	AW511366	AW511366 hd45h03.x
C 33	379.8	59.8	664	5	BA481837	BA481837 DKFZp886M
C 34	375.2	59.1	556	1	AU158595	AU158595 AU158595
C 35	369.6	58.2	865	6	CB231128	CB231128 AGENCOURT
C 36	365.8	57.6	470	1	AI074704	AI074704 ox83d05.s
C 37	364.8	57.4	679	9	AG076758	AG076758 Pan trogl
C 38	364.2	57.4	415	1	AI128496	AI128496 qc61e08.x
C 39	360.6	56.8	777	9	AG030228	AG030228 Pan trogl
C 40	359.8	56.7	425	1	AI570707	AI570707 tm79g09.x
C 41	359.8	56.7	431	1	AA552941	AA552941 nk61a10.s
C 42	359.2	56.6	490	1	AA426511	AA426511 zw02e05.x
C 43	359.2	56.6	619	9	AG133542	AG133542 Pan trogl
C 44	359.2	56.6	674	9	AG091649	AG091649 Pan trogl
C 45	355.4	56.0	446	6	CB069106	CB069106 is13a01.x

ALIGNMENTS

RESULT 1
LOCUS BC030968 1160 bp mRNA linear HTC 19-NOV-2003
DEFINITION Homo sapiens cDNA clone IMAGE:4724433, with apparent retained intron.
ACCESSION BC030968
VERSION BC030968.1 GI:22658419
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1160)
AUTHORS Klausner, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Xlauner, R.F., Collins, F.S., Wagner, L., Shermen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Donald, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prance, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzyzanski, M.I., Skalska, U., Smal, M.A., Schnerk, A., Schein, J.E., Jones, S.J. and Marra, M.A.
GENERATION AND INITIAL ANALYSIS OF MORE THAN 15,000 FULL-LENGTH HUMAN AND MOUSE cDNA SEQUENCES
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
MEDLINE 12477932
PUBMED 12477932
REFERENCE 2 (bases 1 to 1160)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (03-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONTECH

cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: <http://image.lnlnl.gov>
Series: IRAL Plate: 41 Row: m Column: 10
This clone has the following problem: retained intron.

FEATURES

Location/Qualifiers
1. .1160
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4724433"
/tissue_type="Placenta"
/clone_lib="NIH_MGC_79"
/lab_hosts="DH10B"
/note="Vector: pDNR-LIB"

ORIGIN

Query Match 78.6%; Score 498.8; DB 3; Length 1160;
Best Local Similarity 92.6%; Pred. No. 3.4e-133; Indels 0; Gaps 0;
Matches 524; Conservative 0; Mismatches 42;

70 CTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCCCTGGACCGG 129
Db CTACAAATGGAGGCCAAGATGACGTCCAGACTAAGATCTACCGCAGACCCCTGGACCGG 641
71 CCTGCTAGCCCATGCTCGATGTTAATGACTTGAAGGCAACCCCTCCCGAGGAATCTCA 189
Db CCTGCTAGCCCATGCTGATGTTAATGACTAAGGCAACCCCTCCCGAGGAATCTCA 701
72 ACTCAGAACCCCTACTATGCCCAATTCAGCGGGAAGCAGTGTAGAGCGTCTATCAGCCA 249
Db GCTGACAACTCTACTACGCCCAATTCAGCAGGAAGCAGTGTAGAGCGTCTATCAGCCA 761
73 ACCTCCCAACAGACATTTGGGTTTTCTGTTGAGAGGGGGAGCTGAGAGACAGACTAGC 309
Db ACCTCCCAACAGACATTTAGGTTTCTGTTGAGTGGGGAGCTGAGAGACAGACTAGC 821
74 TGGATTTCTAGGCAACAGAGATCCCTAGACCTAGCTGGGAAGGTGACTGCAATCCACC 369
Db TGGATTTCTAGGCTGACTAAGATCCCTAAGCCTAGCTGGGAAGGTGACCAATCCACC 881
75 TCTAAACATGGGCTTGCACTTAGCTCACACCCGACCAATCAGAGAGCTCCTATAAATG 429
Db TTTAAACACGGGCTTGGCACTTAGCTCACACCTGACCAATCAGAGAGCTCCTATAAATG 941
76 CTAATTAGGCAAAATAGAGGTAAAGAAATAGCAATCATCTATTGCTGAGAGCAG 489
Db CTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCAATCATCTATTGCTGAGAGCAG 1001
77 CGGAGGCAACAGATCGGATATAACCCAGGCAATTCGAGCCGCAACGCAACCCCT 549
Db CAGGAGGCAACATGATCGGGATATAACCCAGGCTTTGAGCCGCAACGCAACCCCT 1061
78 TTGGGTCCCTCCCTTTGATGGGCGCTCTGTTTTCACTCTATTTCATCTATTAAATCT 609
Db TTGGGTCCCTCCCTTTGATGGGAGCTCTGTTTTCATGCTATTTCATCTATTAAATCT 1121
79 TCGAACTGAAAAAATAAAAAA 635
Db TCGAACTACAAAAAATAAAAAA 1147

RESULT 2
CR605851
LOCUS CR605851 2748 bp mRNA linear HTC 21-JUL-2004

DEFINITION

full-length cDNA clone CSODE012YJ24 of Placenta of Homo sapiens (human).

ACCESSION
CR605851
VERSION
CR605851.1 GI:50486658
KEYWORDS
HTC; CNSLT_cDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1. (bases 1 to 2748)
AUTHORS
Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
REMARK
Contact : Feng Liang Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600
Paraday Avenue
2 (bases 1 to 2748)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.

FEATURES

Location/Qualifiers
1. .2748
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODE012YJ24"
/tissue_type="Placenta"
/plasmid="pCMVSPORT_6"

ORIGIN

Query Match 77.6%; Score 492.8; DB 3; Length 2748;
Best Local Similarity 90.9%; Pred. No. 2.3e-131;
Matches 540; Conservative 0; Mismatches 42; Indels 12; Gaps 1;

1 CCTGTATCTTTAACTCTCTGTTAAGTTTGTCTTCTCCAGAAATCAAACTGTAAACTTA 60
Db CCTGTATCTTTAACTCTCTGTTAAGTTTGTCTTCTCCAGAAATCGAAGCTGTAAACTTA 2226
61 CAAATGTTTCTTCAATGGAGCACCAAGATGGAGTCCATGACTAAGATCTACCGCAGACCC 120
Db CAAATGTTTCTTCAATGGAGCACCAAGATGGAGTCCATGACTAAGATCTACCGCAGACCC 2274
121 CTGGACCGGCTCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCCCGAG 180
Db CTGGACCGGCTCTAGCCCATGCTCCGATGTTAATGACATTGAAGGCACCCCTCTCTGAG 2334
181 GAAATCTCAACTCACAACCCCTACTATGCCCAATTCAGCGGGAAGCAGTTAGACGGT 240
Db GAAATCTCACTGTCACAACCTCTACTACGCCCAATTCAGCAGGAAGCAGTTAGACGGT 2394
241 CATCAGCCAACTCCCAACAGACCTTGGGTTTTCTGTTGAGAGGGGGAGCTAGAGAC 300
Db CGTCGGCCAACTCCCAACAGACCTTGGGTTTTCTGTTGAGATGGGGAGCTAGAGAC 2454
301 AGACTAGCTGGATTTCTTAGGCCAACAGAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 360
Db AGACTAGCTGGATTTCTTAGGCCAACAGAGAAATCCCTAAGCCTAGCTGGGAAGGTGACT 2514
361 GCATCCACCTCTAAACATATGGGCTTGCACCTTAGCTTACACCCGACCAATCAGAGAGCTC 420
Db ACATCCACCTTTAAACACAGGGCTTGCACCTTAGCTTACACCTGACCAATCAGAGAGCTC 2574
421 ACTAAAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCTG 480
Db ACTAAAAATGCTAATTAGGCAAAAGCAGGAGGTAAAGAAATAGCAATCATCTATTGCTG 2634

```
QY 481 AGAGCACAGCGGAGGACCAAGGATCGGGATATAAACCAGCAATTCGAGCCGCAACGG 540
Db 2635 AGAGCACAGCAGGAGGACCAATGATCGGGATATAAACCAGCAATTCGAGCCGCAACGG 2694
QY 541 CAACCCCTTTGGGTCCCTCCCTTTGATGGGGCTCTGTTTCACTCTATTT 594
Db 2695 CAACCCCTTTGGGTCCCTCCCTTTGATGGGGCTCTGTTTCACTCTATTT 2748

RESULT 3
CR617248 2749 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CSODI022YJ18 of Placenta Cot 25-normalized
DEFINITION of Homo sapiens (human).
ACCESSION CR617248.1 GI:50498055
VERSION HTC; CNSLT cDNA.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2749)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
REFERENCE 2 (bases 1 to 2749)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES             Location/Qualifiers
     source
       1..2749
         /organism="Homo sapiens"
         /mol_type="mRNA"
         /db_xref="taxon:9606"
         /clone="CSODI022YJ18"
         /tissue_type="Placenta"
         /plasmid="pCMVSPORT_6"
ORIGIN

Query Match       77.0%; Score 488.8; DB 3; Length 2749;
Best Local Similarity 90.8%; Pred. No. 3.3e-130;
Matches 536; Conservative 0; Mismatches 42; Indels 12; Gaps 1;

QY 1 CCCTGTATCTTTAACTCCCTTGTAGTTGTCTCTCCAGATCAAACTGTAAACTA 60
Db 2172 CCCTGTATCTTTAACTCCCTTGTAGTTGTCTCTCCAGATCGAAGCTGTAAACTA 2231
QY 61 CAAATTGTTCTTCAAATGGAGCACAGATGAGTGCCATGACTAAGATCCACCGTGACCC 120
Db 2232 -----CAAATGGAGCCCAAGATGAGTCCAGACTAAGATCTACCGCAGACCC 2279

QY 121 CTGACCGGCTGTAGCCCATGCTCCGATGTTTAATGACATTTGAAGCACCCCTCCCGAG 180
Db 2280 CTGACCGGCTGTAGCCCATGCTCCGATGTTTAATGACATCAAAAGGCACCCCTCCCTGAG 2339

QY 181 GAAATCTCAACTGCACAACCCCTACTATGCCCAATTCAGGGGAGCAGTTAGACGGT 240
Db 2340 GAAATCTCAGTGCACAACCTCTACTACGCCCAATTCAGCAGGAGCAGTTAGACGGT 2399

QY 241 CATCAGCAACCTCCCAACAGCACTTTGGGTTTTCTCTGTTGAGAGGGGAGCTGAGAGAC 300
Db 2400 CGTCGGCAACCTCCCAACAGCACTTAGGTTTTCTCTGTTGAGATGGGGAGCTGAGAGAC 2459
```

```
QY 301 AGACTAGCTGGATTTCTTAGGCCAACGAGAGATCCCTTAGCCTAGCTGGGAAGGTGACT 360
Db 2460 AGACTAGCTGGATTTCTTAGGCCAACGAGAGATCCCTTAGCCTAGCTGGGAAGGTGACC 2519
QY 361 GCATCCACCTCTAAACATGGGGCTTGCACTTAGCTCACAACCCGACCAATCAGAGACTC 420
Db 2520 ACATCCACCTTTAAACACGCGGCTTGCACTTAGCTCACAACCTGACCAATCAGAGACTC 2579
QY 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCTG 480
Db 2580 ACTAAATGCTAATTAGGCAAAAGACAGAGAGGTAAAGAAATAGCAATCATCTATTGCTG 2639
QY 481 AGAGCACAGCGGAGGACCAAGGATCGGGATATAAACCAGGCATTCGAGCCGCAACGG 540
Db 2640 AGAGCACAGCAGGAGGACCAATGATCGGGATATAAACCAGGCATTCGAGCCGCAACGG 2699
QY 541 CAACCCCTTTGGGTCCCTCCCTTTGATGGGGCTCTGTTTCACTCT 590
Db 2700 CAACCCCTTTGGGTCCCTCCCTTTGATGGGGCTCTGTTTCACTGCT 2749

RESULT 4
CR622175 2500 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CSODI051YM13 of Placenta Cot 25-normalized
DEFINITION of Homo sapiens (human).
ACCESSION CR622175.1 GI:50502982
VERSION HTC; CNSLT cDNA.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2500)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
REFERENCE 2 (bases 1 to 2500)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES             Location/Qualifiers
     source
       1..2500
         /organism="Homo sapiens"
         /mol_type="mRNA"
         /db_xref="taxon:9606"
         /clone="CSODI051YM13"
         /tissue_type="Placenta"
         /plasmid="pCMVSPORT_6"
ORIGIN

Query Match       76.9%; Score 488; DB 3; Length 2500;
Best Local Similarity 91.1%; Pred. No. 5.5e-130;
Matches 534; Conservative 0; Mismatches 40; Indels 12; Gaps 1;

QY 1 CCCTGTATCTTTAACTCCCTTGTAGTTGTCTCTCCAGATCAAACTGTAAACTA 60
Db 1925 CCCTGTATCTTTAACTCCCTTGTAGTTGTCTCTCCAGATCGAAGCTGTAAACTA 1984
QY 61 CAAATTGTTCTTCAAATGGAGCACAGATGAGTGCCATGACTAAGATCCACCGTGACCC 120
Db 1985 -----CAAATGGAGCCCAAGATGAGTCCAGACTAAGATCTACCGCAGACCC 2032
```

```
QY 121 CTGGACGGGCTGCTAGCCATGCTCCGATGTTAATGACATTGAAGGACCCCTCCCGAG 180
Db 2033 CTGGACGGGCTGCTAGCCACGATCTGATGTTAATGACATTGAAGGACCCCTCTTGAG 2092
QY 181 GAAATCTCAACTGCACAACCCCTACTATGCCCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
Db 2093 GAAATCTCAGCTGCACAACCTCTACTACGCCCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2152
QY 241 CATCAGCAACCTCCCAACAGACACTTGGGTTTTCTCTGTTGAGAGGGGGGACTGAGAGAC 300
Db 2153 CGTCGGCAACCTCCCAACAGACACTTGGGTTTTCTCTGTTGAGATGGGGGACTGAGAGAC 2212
QY 301 AGGACTAGCTGGATTTCTTAGGCCCAAGAGATCCCTAAGCCTAGCTAGCTGGGAAGGTGACT 360
Db 2213 AGGACTAGCTGGATTTCTTAGGGCTGACTAAGATCCCTAAGCCTAGCTGGGAAGGTGACC 2272
QY 361 GCATCCACCTCTAAACATATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
Db 2273 ACATCCACCTTTAAACACAGCGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTC 2332
QY 421 ACTAAATGCTAATAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 480
Db 2333 ACTAAATGCTAATAGGCAAAAGCAGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 2392
QY 481 AGACACAGCGGGAGGACAGGATCGGGATATAAACCCAGGCAATTCGAGCCGCAACGG 540
Db 2393 AGACACAGCAGGAGGACAGATGATCGGATATAAACCCAGGCTTCGAGCCGCAACGG 2452
QY 541 CAACCCCTTTGGTCCCTCCCTTTGATGGGCGCTCTGTTTCA 586
Db 2453 CAACCCCTTTGGTCCCTCCCTTTGATGGGAGCTCTGTTTCA 2498
```

```
RESULT 5
BI087886 771 bp mRNA linear EST 20-JUN-2001
LOCUS 602852690F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4993894 5',
DEFINITION mRNA sequence.
ACCESSION BI087886.1 GI:14506216
VERSION BI087886.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL1015 row: d column: 23
High quality sequence stop: 762.
Location/Qualifiers
1. .771
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4993894"
/cell_line="MGC36"
/lab_host="DH10B"
/clone_lib="NIH_MGC_10"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5 kb. Library prepared by Life
Technologies."
```

FEATURES

source

```
ORIGIN
Query Match 76.5%; Score 485.6; DB 4; Length 771;
Best Local Similarity 89.9%; Pred. No. 2.1e-129;
Matches 570; Conservative 0; Mismatches 49; Indels 15; Gaps 4;
QY 1 CCCTGTATCTTTAACTCTCTTGGTTAGTTTCTCTTCCAGAAATCAAAACCTGTAATACTA 60
Db 151 CCCTGTATCTTTAACTCTCTTGGTTAGTTTCTCTTCCAGAAATCGAAGCTGTAAATACTA 210
QY 61 CAATTTGTTCTTCAAAATGGAGCACAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
Db 211 -----CAATTTGAGGCCCAAGATGCAGTCCAAAGACTAAGATCTACCGCAGACCC 258
QY 121 CTGGACGGGCTGCTAGCCATGCTCCGATGTTAATGACATTGAAGGACCCCTCCCGAG 180
Db 259 CTGGACGGGCTGCTAGCCACGATCTGATGTTAATGACATTGAAGGACCCCTCCCGAG 318
QY 181 GAAATCTCAACTGCACAACCCCTACTATGCCCAATTCAGCGGGAAGCAGTTAGAGCGGT 240
Db 319 GAAATCTCAGCTGCACAACCTCTACTACGCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 378
QY 241 CATCAGCAACCTCCCAACAGCAGCTTGGGTTTTCTCTGTTGAGAGGGGGACTGAGAGA 299
Db 379 CGTC-GGCAACCTCCCAACAGCAGCTTGGGTTTTCTCTGTTGAGATGGGGACTGAGAGA 437
QY 300 CAGGACTAGCTGGATTTCTTAGGCCAACGAAGAATCCCTAGCCTAGCTGGGAAGGTGAC 359
Db 438 CAGGACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGGGAAGGTGAC 497
QY 360 TGCATCCACCTCTAAACATGGGGCTTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCT 419
Db 498 CACATCCACCTTTAAACACGGGGCTTGCNACTTAGCTCACACCTGACCAATCAGAGAGCT 557
QY 420 CACTAAATGCTAATTAGGCAAAATAGGAGGTA-AAGAAATAGCAATCATCTATTGCC 478
Db 558 CACTAAATGCTAATTAGGCAAAAGCAGAGGTACAAGAAATAGCAATCATCTATTGCC 617
QY 479 TGAGAGCAGCGGGAGGAGGACAGGATCGGATATAAACCCAGGCAATTCGAGCCGCAAC 538
Db 618 TGAGAGCAGCAGCAGGAGGAGCAATGATCGGGATATTTAAACCAAGTCTTCGAGCCGCAAC 677
QY 539 GGCAACCCCTTTGGGTCCTCCCTCTTGTATGGGCGCTCTGTTTCACTATTTCCT 598
Db 678 GGCAACCCCTTTGGTCCCTCTTGTATGGGAGCTCTGTTTCACTATTTCCT 737
QY 599 CTATTAAATCTTCAACTGAAAAAATAAAAAA 632
Db 738 CTATTAAATCTTCAACTGAAAAAATAAAAAA 771
RESULT 6
CR625046 2748 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CS0D1044YK06 of Placenta Cot 25-normalized
DEFINITION of Homo sapiens (human).
ACCESSION CR625046.1 GI:50505853
VERSION CR625046.1
KEYWORDS HTC; CNSLUT_cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 2748)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
JOURNAL
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue
2 (bases 1 to 2748)
Genoscope.
Direct Submission
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
REFERENCE
AUTHORS
TITLE
```


LOCUS BX365066 1071 bp mRNA linear EST 08-APR-2004
 DEFINITION BX365066 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
 clone CS0D1046VA18 3-PRIME, mRNA sequence.
 ACCESSION BX365066
 VERSION BX365066.2 GI:46304105
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1071)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 On May 5, 2003 this sequence version replaced gi:30374869.
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 4215.r
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?s=CS1A1012ZC1ONP1&c=4215.r.
 FEATURES
 Location/Qualifiers
 1..1071
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0D1046VA18"
 /tissue_type="PLACENTA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."
 ORIGIN
 Query Match 73.8%; Score 468.6; DB 5; Length 1071;
 Best Local Similarity 89.9%; Pred. No. 1.9e-124;
 Matches 525; Conservative 3; Mismatches 43; Indels 13; Gaps 2;
 1 CCTGTATCTTTAACTCCCTTGTAGTTGTTCTCTCCAGAACTCAAACTGTAATACTA 60
 571 CCCTGTATCTTTAACTCCCTTGTAGTTGTTCTCTCCAGAACTCAAACTGTAATACTA 512
 61 CAATATTGTTCTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
 511 -----CAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 464
 121 CTGACCGGCTGCTAGCCATGCTCCGATGTTTAATGACATTGAAGCACCCTCCCGAG 180
 463 CTGACCGGCTGCTAGCCATGCTCCGATGTTTAATGACATTGAAGCACCCTCCCGAG 404
 181 GAATCTCAACTGACACACCCCTACTATGCCCCAAATTCAGCGGAAGCAGTTAGACGGT 240
 403 GAATCTCAACTGACACACCCCTACTATGCCCCAAATTCAGCGGAAGCAGTTAGACGGT 344
 241 CATCAGCACAACCTCCCAACAGCACTTGGGTTTTCTGTTGAGAGGGGGAGCTGAGAGAC 300
 343 CGTCGGCAACCTCCCAACAGCACTTGGGTTTTCTGTTGAGAGGGGGAGCTGAGAGAC 284
 301 AGGACTAGCTGGATTTCTAGGCCAACGAAGAATCCCTAGCCTAGCTGGGAAGTGAAT 360
 283 AGGACTAGCTGGATTTCTAGGCCAACGAAGAATCCCTAGCCTAGCTGGGAAGTGAAT 224
 361 GCATCCACCTCAACATGGGGCTTGCAACTTAGCTCACACCGCAGCAATCAGAGAGCTC 420
 223 ACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCGCAGCAATCAGAGAGCTC 164

QY 421 ACTAAATGCTTAATAGCAAAATAGCAGGTAAGAAATAGCAATCATCTATTGCGCTG 480
 Db 163 ACTAAATGCTTAATAGCAAAATAGCAGGTAAGAAATAGCAATCATCTATTGCGCTG 104
 QY 481 AGAGCACAGCGGAGGAGCAAGGATCGGATATAAACCAGGCATTTCGAGCGGCAACGG 540
 Db 103 AGAGCACAGCGGAGGAGCAAGGATCGGATATAAACCAGGCATTTCGAGCGGCAACGG 44
 QY 541 CAACCCCTTTGGGTCCTCCCTTGTATGGGCGCTCTGTTTT 584
 Db 43 CAACCCCTTTGGTNCCTCCCTCCCTTTGTATGGGCGCTCTGTTTT 1
 RESULT 9
 BX357208/c
 LOCUS
 DEFINITION BX357208 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
 clone CS0D1022YJ18 3-PRIME, mRNA sequence.
 ACCESSION BX357208
 VERSION BX357208.2 GI:4630595
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 758)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 On May 5, 2003 this sequence version replaced gi:30376125.
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 4215.r
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?s=CS0D1022DE09NP1&c=4215.r.
 FEATURES
 Location/Qualifiers
 1..758
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0D1022YJ18"
 /tissue_type="PLACENTA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."
 ORIGIN
 Query Match 73.5%; Score 466.6; DB 5; Length 758;
 Best Local Similarity 88.8%; Pred. No. 6.8e-124;
 Matches 509; Conservative 9; Mismatches 43; Indels 12; Gaps 1;
 1 CCCTGTATCTTTAACTCCCTTGTAGTTGTTCTCTCCAGAACTCAAACTGTAATACTA 60
 576 CCCTGTATCTTTAACTCCCTTGTAGTTGTTCTCTCCAGAACTCAAACTGTAATACTA 517
 61 CAATATTGTTCTTCAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
 516 -----CAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 469
 121 CTGACCGGCTGCTAGCCATGCTCCGATGTTTAATGACATTGAAGCACCCTCCCGAG 180
 468 CTGACCGGCTGCTAGCCATGCTCCGATGTTTAATGACATTGAAGCACCCTCCCGAG 409
 181 GAATCTCAACTGACACAACCCCTTACTATGCCCCAAATTCAGCGGAAGCAGTTAGACGGT 240

```
Db 408 GAAATCTCAGCTGCAAACTCTTACTAGCCCAATTCACAGGAAGCACTTGAAGCGGT 349
Qy 241 CATCAGCAACCTCCCAACAGCACCTTGGGTTTTCTGTTGAGAGGGGGAGCTGAGAGAC 300
Db 348 CGTCGGCCAACTCCCAACAGCACCTTGGGTTTTCTGTTGAGATGGGGAGCTGAGAGAC 289
Qy 301 AGACTAGCTGGATTTCTTAGGCAACAGCAATCCCTAAGCCTAGCTAGTGGGAAGGTGACT 360
Db 288 AGACTAGCTGGATTTCTTAGGCTGACTAAGAACTCCCTAAGCCTAGCTAGTGGGAAGGTGACC 229
Qy 361 GCATCCACCTCTAAACATGGGCTTGCACCTAGCTCAGACCCGAGCAATCAGAGAGCTC 420
Db 228 ACATCCACCTTTAAACAGGGCTTGCACCTAGCTCAGACCTGACCAATCAGAGAGCTC 169
Qy 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 480
Db 168 ACTAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 109
Qy 481 AGAGCAGCGGGAGGAGCAAGATCGGGATATAAACCAGGCAATTCGAGCCGGCAACGG 540
Db 108 AGAGCAGCAGGAGGAGCAATGATCGGGATATAAACCAGGCAATTCGAGCCGGCAACGG 49
Qy 541 CAACCCCTTTGGTCCCTCCCTTTCTGATGGG 573
Db 48 CMACCCCTTGGTCCCTCCCTTTGATGGG 16

RESULT 10
LOCUS CR613169 2716 bp mRNA linear HTC 21-JUL-2004
DEFINITION full-length cDNA clone CS0DE013Y120 of Placenta of Homo sapiens
(human).
ACCESSION CR613169.1 GI:50493976
VERSION HPC; CDSLT_cDNA.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2716)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 2716)
Genoscope.
Direct Submission
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source Location/Qualifiers
1..2716
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE013Y120"
/tissue_type="Placenta"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 73.2%; Score 464.6; DB 3; Length 2716;
Best Local Similarity 90.9%; Pred. No. 3.4e-123;
Matches 510; Conservative 0; Mismatches 39; Indels 12; Gaps 1;
Qy 1 CCCTGTATCTTTAACTCTTGTGTTAGTGTCTCTCCAGAATCAAAACTGTAATAACTA 60
```

```
Db 2168 CCCTGTATCTTTAACTCTTGTGTTAGTGTCTCTCCAGAATCGAAGCTGTAAACTA 2227
Qy 61 CAAATTTGTTCTTCAAAATGGAGCACCAAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
Db 2228 -----CAAAATGGAGCCCAAGATGCACTCCCAAGACTAAGATCTACCGCAGACCC 2275
Qy 121 CTGGACGGGCTGCTAGCCATGCTCCGATGTTAATGACATTTGAAGGCACCCCTCCCGAG 180
Db 2276 CTGGACGGGCTGCTAGCCATGCTGATGTTAATGACATCAAGAGCACCCTCTCTGAG 2335
Qy 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCCAAATTCAGCGGGAAAGCAGTTAGAGCGGT 240
Db 2336 GAAATCTCAGTGCACAAACCTCTACTACGCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 2395
Qy 241 CATCAGCCCAACCTCCCAACAGCACTTGGGTTTTCTGTTGAGAGGGGGAGCTGAGAGAC 300
Db 2396 CGTCGGCCAACTCCCAACAGCACTTGGGTTTTCTGTTGAGATGGGGAGCTGAGAGAC 2455
Qy 301 AGACTAGCTGGATTTCTTAGGCAACAGCAATCCCTAAGCCTAGCTAGTGGGAAGGTGACT 360
Db 2456 AGACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGGGAAGGTGACC 2515
Qy 361 GCATCCACCTCTAAACATGGGCTTGCACCTAGCTCAGACCCGAGCAATCAGAGAGCTC 420
Db 2516 ACATCCACCTTTAAACAGGGCTTGCACCTAGCTCAGACCTGACCAATCAGAGAGCTC 2575
Qy 421 ACTAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 480
Db 2576 ACTAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCAATCATCTATTGCGCTG 2635
Qy 481 AGAGCAGCGGGAGGAGCAAGATCGGGATATAAACCAGGCAATTCGAGCCGGCAACGG 540
Db 2636 AGAGCAGCAGGAGGAGCAATGATCGGGATATAAACCAGGCAATTCGAGCCGGCAACGG 2695
Qy 541 CAACCCCTTTGGTCCCTCCCTC 561
Db 2696 CAACCCCTTTGGGTCCTC 2716

RESULT 11
LOCUS BX337769/c 998 bp mRNA linear EST 07-APR-2004
DEFINITION BX337769 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1051YM13 3-PRIME, mRNA sequence.
ACCESSION BX337769
VERSION BX337769.2 GI:46272079
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 998)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 2, 2003 this sequence version replaced gi:30337641.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
4215.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0D1051AG07NP1&c=4215.r.
FEATURES
source Location/Qualifiers
1..998
/mol_type="mRNA"
```

/db_xref="taxon:9606"
/clone="CS0D1051YML3"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 71.5%; Score 454.2; DB 5; Length 998;
Best Local Similarity 89.1%; Pred. No. 2.9e-120;
Matches 528; Conservative 2; Mismatches 47; Indels 15; Gaps 3;
Qy 1 CCTGTATCTTTAACTCTCTTGAAGTTTGTCTCTTCAGATCAAAACTGTAAACTA 60
Db 579 CCCTGTATCTTTAACTCTCTTGAAGTTTGTCTCTTCAGATCAAAACTGTAAACTA 520
Qy 61 CAAATTTCTTCAATGGAGCACCAGATCGAGTCCATGACTAAGATCCACCGTGCACCC 120
Db 519 -----CAAATGGAGCACCAGATCGAGTCCATGACTAAGATCCACCGTGCACCC 472
Qy 121 CTGACCGCGCTGCTAGCCATGCTCCGATGTTAATGACATGAA-GGACACCCCTCCCGA 179
Db 471 CTGACCGCGCTGCTAGCCATGCTGATGTTAATGACATGAAAGCGCACCCCTCTGTA 412
Qy 180 GGAATCTCACTGCACACACCCCTACTATGCCCCCAATTCAGCGGAGCAGTTAGACGG 239
Db 411 GGAATCTCACTGCACACACCCCTACTACGCCCAATTCAGCGGAGCAGTTAGACGG 352
Qy 240 TCATCAGCAACCTCCCAACAGCAGCTTGGGTTTCTCTGTTGAGAGGGGGAGCTGAGAGA 299
Db 351 TCGTCGGCCACCTCCCAACAGCAGCTTAGGTTTCTCTGTTGAGAGGGGGAGCTGAGAGA 292
Qy 300 CAGGACTAGCTGGAATTTCTAGGCAACAGAGATCCCTAGCTAGCTGGGAGGTGAC 359
Db 291 CAGGACTAGCTGGAATTTCTAGGCTGACTAAGATCCCTAGCTAGCTGGGAGGTGAC 232
Qy 360 TGCATCCACCTTAACATAGGGGTTGCACTTAGCTCACACCGCAGCAATCAGAGAGCT 419
Db 231 CACATCCACCTTTAAACAGGGGTTGCACTTAGCTCACACCTGAGCTCAGAGAGCT 172
Qy 420 CACTAAATGCTAATTAGGCAAAATAGAGGTTAAAGAAATAGCAATCATCTATTGCTCT 479
Db 171 CACTAAATGCTAATTAGGCAAAAGACAGAGGTTAAAGAAATAGCAATCATCTATTGCTCT 112
Qy 480 GAGAGCAGCGGGAGGAGCAGAGATCGGGATATAAACCCAGCATTCGAGCGGCAACG 539
Db 111 GAGAGCAGCAGGAGGAGCAGATGATCGGGATATAAACCCAGCTCTCGAGCGGCAACG 52
Qy 540 GCAACCCCTTTGGGTCCC--CTCCCTTTGATGGGCGCTCTGTTTCA 586
Db 51 GCAACCCCTTTGGGTCCCNCCCTTTGTTANTGGAGGCTCTGTTTCA 3

RESULT 12

LOCUS BX378303/c
DEFINITION BX378303 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1051YH07 3-PRIME, mRNA sequence.
BX378303.2 GI:46557492

ACCESSION

BX378303

VERSION

BX378303.2

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 1058)

AUTHORS

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished (2001)

COMMENT

On May 8, 2003 this sequence version replaced gi:30439129.
Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
4215.r

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?cs=CS0D1018CD04NP1&c=4215.r.

Location/Qualifiers

1..1058

FEATURES

source

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0D1018YH07"

/tissue_type="PLACENTA COT 25-NORMALIZED"

/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

sites of the pCMVSPORT 6 vector. Library was normalized."

AG121669 689 bp DNA linear GSS 04-NOV-2001

RESULT 13

AG121669/c

LOCUS

```

DEFINITION Pan troglodytes DNA, clone: PTB-130M15.F, genomic survey sequence.
ACCESSION AG121669
VERSION AG121669.1 GI:16650834
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 1 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
AUTHORS 2 (bases 1 to 689)
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Aaso Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpses@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
FEATURES
source
Location/Qualifiers
1. .689
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-130M15.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
Query Match 71.0%; Score 450.8; DB 9; Length 689;
Best Local Similarity 88.3%; Pred. No. 2.6e-119;
Matches 515; Conservative 0; Mismatches 58; Indels 10; Gaps 2;
QY 1 CCCTGTATCTTTAACTCCTCTTGTAGTTTGTCTTCCAGAAATCAAACTGTAATACTA 60
DB 666 CCNIGGATCTTTAAGCTCTTGTAGCTTGTCTTCCAGAAATCAAGCTGTAATACTA 607
QY 61 CAAATGTTCTTCAAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGGACCC 120
DB 606 TAATAGTTTCTTCAAATGGAGCACCATTATGACGTCCTGACTAAGATCTACCATGGACCC 547
QY 121 CTGGACGGCTGCTAGCCCATGCTCGGATGTTAATGACATTCGAAGCACCCTCCCGAG 180
DB 546 CTGGACGAGCTGCTAGCCCATGCTCTGATGTTAATGACATTCGAAGCACCCTCCCTGAG 487
QY 181 GAAATCTCAACTGCAC-APCCCTACTATGSCCAATTCAGCGGAGCAGTTAGAGCGG 239
DB 486 GAAATCTCAACTGCACAAACCTTACTACATCCAGTTCAGCAGGAGCAGTTAGAGCGG 427
QY 240 TCATCAGCAACCTCCCCAAACAGCAGCTTGGGTTTCTGTTGAGAGGGGGGACTGAGAGA 299
DB 426 TCGTCGGCAACCTCCCCANTGSCACTTGGGTTTCTGTTGAGAGGGGGGACTGAGAGA 367
QY 300 CAGACTAGCTGATTTCTTAGGCCAAACGAAGAATCCCTAAGCCTAGCTGGGAAGGTGAC 359
DB 366 CAGGACTATCTGATTTCTTAGGCCAAACCTTACTACATCCAGTTCAGCAGGAGCAGTTAGAGCGG 307
QY 360 TGCATCCACTCTAAACATGGGCTTGCAACTTAGCTTACACCCGACCCACCATC----- 411
DB 306 CACATTCACCTTTTAAACACAGGGCTTGCAACTTAGCTTACACCCGACCCGACCATCAGGTAGTA 247

```

```

QY 412 -AGAGAGCTCACTAAAATGCTAATTAGGCAGAAAATAGGAGGTAAAGAAATAGCAATCAT 470
DB 246 AAGAGGGCTCACTAAAATGCTAATTAGGCAGAAAATAGGAGGTAAAGAAATAGCAATCAT 187
QY 471 CTATTCCTCTGAGAGCAGCGGAGGACCAAGGATCGGATATATAAACCAGGATTCGAG 530
DB 186 TTATTCCTCTGAGAGTACAGCGGAGGACCAAGGATCGGATATATAAACCAGGATTCGAG 127
QY 531 CCGGCAACGGCACCCCTTTGGTCCCTCCCTCTTGTATGGG 573
DB 126 CCACCAATGCTACCCCTTTGGTCCCTCCCTCTTGTATGG 84

RESULT 14
LOCUS AA781423/c
DEFINITION
AA781423 494 bp mRNA linear EST 31-DEC-1998
aj26c03.s1 Soares testis NHT Homo sapiens cDNA clone 1391428 3'
similar to contains PRR7.t1 PRR7 repetitive element ;, mRNA
sequence.
ACCESSION AA781423
VERSION AA781423
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 494)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1645 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 475.
FEATURES
source
Location/Qualifiers
1. .494
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="1391428"
/sex="male"
/lab_host="DH10B"
/clone_lib="Soares testis NHT"
/notes="Vector: pRT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5].
TGTTCACCATCTGAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3'}.
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pRT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
ORIGIN
Query Match 68.8%; Score 437; DB 1; Length 494;
Best Local Similarity 92.7%; Pred. No. 2.4e-115;
Matches 458; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY 131 CTGCTAGCCCATGCTCCGATGTTAATGACATGAAGGACCCCTCCCGAGGAATCTCAA 190
DB 494 CTGCTAGCCCATGCTCCGATGTTAATGACATGAAGGACCCCTCCCGAGGAATCTCAG 435

```

double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

QY	191	CTGCACAAACCCCTATATGCCCCCAATTCAGCGGGAAGAGCTTAGAGCCGCTCATCAGCCAA	250
Db	434	CTGCACAAACCCCTATATGCCCCCAATTCAGCGGGAAGAGCTTAGAGCCGCTCATCAGCCAA	375
QY	251	CTTCCCAACAGACACTTGGGTTTCTCTGTTGAGAGGGGAGCTTGAGAGACAGACTAGCT	310
Db	374	CTTCCCAACAGACACTTGGGTTTCTCTGTTGAGAGGGGAGCTTGAGAGACAGACTAGCT	315
QY	311	GGATTTCCTAGGCAACGAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACTGATCCACCT	370
Db	314	GGATTTCCTAGGCAACGAGAAATCCCTAAGCCCTAGCTGGGAAGGTGACTGATCCACCT	255
QY	371	CTAAACATGGGGCTTGCACTTAGCTCACACCCGCAATCAGAGAGCTCACTAAATGC	430
Db	254	TTAAACACGGGGCTTGCACTTAGCTCACACCCGCAATCAGAGAGCTCACTAAATGC	195
QY	431	TAAATAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTTCCTGAGAGACAGC	490
Db	194	TAACTAGGCAAAAGACAGAGGTAAAGAAATAGCAATCATCTATTTCCTGAGAGACAGC	135
QY	491	GGGAGGACAAAGATCGGGATATAAACCAGGCAATTCGAGCCGGCAACGGCAACCCCTT	550
Db	134	AGGAGGACAAATGATCGGGATATAAACCAGGCAATTCGAGCCGGCAACGGCAACCCCTT	75
QY	551	TGGTCCCTCCCTTCTGTTGGGGCTCTGTTTCACTCTATTTCCTCTATTAAATCTT	610
Db	74	TGGTCCCTCCCTTCTGTTGGGGCTCTGTTTCACTCTATTTCCTCTATTAAATCTT	15
QY	611	GCAACTGAAAAAAA	624
Db	14	GCAACTGAAAAAAA	1

RESULT 15
LOCUS BX439636/c 1019 bp mRNA linear EST 04-MAY-2004
DEFINITION BX439636 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE012YJ24
3-PRIME, mRNA sequence.
ACCESSION BX439636
VERSION BX439636.2 GI:47000005
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1019)
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 15, 2003 this sequence version replaced gi:30771765.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized. Library was constructed by Life Technologies, a division of Invitrogen.
This sequence belongs to sequence cluster 4215.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?c=CS0DE012DE12NP1&c=4215.r.
Location/Qualifiers
1..1019
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE012YJ24"
/issue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched,

ORIGIN

Query Match 68.5%; Score 435; DB 5; Length 1019;
Best Local Similarity 83.4%; Pred. No. 1.1e-114; Indels 13; Gaps 2;
Matches 494; Conservative 19; Mismatches 66;

QY	1	CCCTGTATCTTTAAACCTCCCTTTGTTTAAAGTTTCTCTTCCAGAAATCAAAAACCTGTAATACTA	60
Db	579	CCCTGTATCTTTAAACCTCCCTTTGTTTAACTTTGTTCTCTTCCAGAAATCGAAGCTGTAATACTA	520
QY	61	CAAAATTTGTTCTTCAAAATGGAGCACACAGATGAGTGCATGATTAAGATCAATCCACCTGGACCC	120
Db	519	-----CAAAATGGAGCCCAAGATGAGTGCATTAAGATCAATCCACCTGGAGACCC	472
QY	121	CTGGACCCGGCTGCTAGCCGATGCTCCGATGCTTAAATGACATTTAAGGACCCCTCCCGAG	180
Db	471	CTGGACCCGGCTGCTAGCCGATGCTGATGTTAATGACATCAAAAGCACCCCTCCTGAG	412
QY	181	GAAATCTCAACTGCAACAAACCCCTACTATGCCCCCAATTTCAGCGGGAAGCAGTTAGAGCGGT	240
Db	411	GAAATCTCAGCTGMAACACCTATATCTAGCCCCCAATTTCAGCAGGAAGCAGTTAGAGCGGT	352
QY	241	CATCAGCAACCTCCCAACAGCACTTGGGTTTCTCTGTTGAGAGGGGAGCTGAGAGAC	300
Db	351	CGTCGGCAACCTCCCAACAGCACTTAGGTTTCTCTGTTGAGATGGGGAGCTGAGAGAC	292
QY	301	AGGACTAGCTGATTTCTTAGGCCCAACGAAGATCCCTAAGCCTAGCTGGGAAGGTGACT	360
Db	291	AGGACTAGCTGATTTCTTAGGCCCAACGAAGATCCCTAAGCCTAGCTGGGAAGGTGACC	232
QY	361	GCATCCACCTTAAACATGGGGCTTGCACCTTAGCTCACACCCGACCAATCAGAGAGCTC	420
Db	231	ACATCCACCTTAAACACAGGGCTTGCACCTTAGCTCACACCTGACCAATCAGAGAGCTK	172
QY	421	ACTAAATCTTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTTCCTG	480
Db	171	ACTAAATCTTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTTCCTG	112
QY	481	AGAGCACAGCGGAGGACAAAGATCGGGATATAAACCAGGCAATTCGAGCCGGCAACGG	540
Db	111	AGAGCACAGCGGAGGACAAAGATCGGGATATAAACCAGGCAATTCGAGCCGGCAACAG	52
QY	541	CAACCCCTTGGGTCCTCCCTTGTATGGCGCTCTGTTTCACTCTAT	592
Db	51	AAACCCYTTT-TTTCYCYCTTTTGTGNTGGNDTTTTTNNNTTATTT	1

RESULT 16
LOCUS AW971553/c 522 bp mRNA linear EST 01-JUN-2000
DEFINITION EST383642 MAGE resequences, MAGU Homo sapiens cDNA, mRNA sequence.
ACCESSION AW971553
VERSION AW971553.1 GI:8161399
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 522)
Hagde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt,J.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and Quackenbush,J.
Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208

FEATURES
source

Email: johnq@tigr.org

Plate: 292

Seq primer: Forward.

Location/Qualifiers

FEATURES

source
1. 522
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="MAGE resequencences, MAGL"
/note="Vector: pBluescriptSkm"

ORIGIN

Query Match 68.1%; Score 432.4; DB 2; Length 522;
Best Local Similarity 91.8%; Pred. No. 5.2e-114;
Matches 479; Conservative 0; Mismatches 41; Indels 2; Gaps 2;
QY 97 ATGACTAAGATCCACCGTGACCCCT-GGACCGGCTGTGTAGCCCATGCTCCGATGTAA 155
Db AGGTTTAAAGATCTTCGGGACCCCTAAACTGGCTGTAGCCCATGTTCTGGTGTAA 463
QY 156 TGACATTGAAGGACCCCTCCGAGGAAA-TCTCAACTGACAAACCCCTACTATGCCCCA 214
Db TGACATCGAAGTCACTCTCTCGAGGAAATTTCAACTGACAAACCCCTATTATGCCCCA 403
QY 215 ATTACGGGGAAGCAGTTAGAGCGGTCACTAGCCNACCTCCCAACAGCACTTTGGCTTTT 274
Db GTTCAGCAGGAAGCAGTTAGAGTGGTCACTAGCCNACCTCCCAACAGCACTTTGGCTTTT 343
QY 275 CTTGTTGAGAGGGGACTCAGAGACAGGACTAGCTGGATTTCCTAGGCCAACGAAGAT 334
Db CTTGTTGAGAGGGGACTCAGAGACAGGACTAGCTGGATTTCCTAGGCCAACGAAGAT 283
QY 335 CCCTAAGCCTTAGTGGGAAGGTGACATGCCATCCACTCTAAACATGGGGCTTGCACCTAG 394
Db CCCTAAGCCTTAGTGGGAAGGTGACCGCATCCACTCTTAAACACGGGGCTTGCACCTAG 223
QY 395 CTCACACCCGACCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAAATAGAGGTAA 454
Db CTCACACCCGACCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAAATAGAGGTAA 163
QY 455 AGAAATAGCCAAATCATCTATTGCTGAGACACAGCGGAGGACAGGATCGGGATATA 514
Db AGAAATAGCCAAATCATCTATTGCTGAGACACAGTGGGAGGACAGGATTCGCAATATA 103
QY 515 AACCCAGGCAATTCAGCGCGCAACCGCAACCCCTTTGGGTCCCTCCCTTTGATGGGC 574
Db AACCCAGGCAATTCAGCGCGCAACCGCAACCGCTTTGGGTCCCTTTCTGATGGGA 43
QY 575 GCTCTGTTTCACTCTATTTCACCTCTATTAAATCTTGCAACT 616
Db GCTCTGTTTCACTCTATTTCACCTCTATTAAATCTTGCAACT 1

RESULT 17

CN645411 870 bp mRNA linear EST 13-MAY-2004
LOCUS ILLUMIGEN MQ 23968 Katze MMSP Macaca mulatta cDNA clone
DEFINITION IBIUW:10227 5' similar to Bases 165 to 770 highly similar to human
Unigene Hs.349001, mRNA sequence.

CN645411 GI:47158854

EST.

Macaca mulatta (rhesus monkey)

ORGANISM

Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Cercopitheciinae; Macaca.

REFERENCE

AUTHORS Katze, M.G., Thomas, M., Korth, M., Iadonato, S.P. and Magnus, C.L.
TITLE Large-scale Rhesus Macaque cDNA Sequencing
JOURNAL Unpublished (2003)
COMMENT Contact: C. Magnus
Illumigen Biosciences Inc.

2203 Airport Way S, Suite 450, Seattle, WA 98134, USA

Tel: 2063780400

Fax: 2063780408

Email: cmagnus@illumigen.com

Sequenced on 2004.02.24. 676 Q20 bases.

PCR Primers

FORWARD: CCTCACTAAAGGGAACAAA

BACKWARD: CACTATAGGGGAATTGGGTA

Insert Length: 870 Std Error: 0.00

Plate: CL00135 row: H column: 02

Seq primer: CCTCACTAAAGGGAACAAA

POLYA=No.

FEATURES

source

Location/Qualifiers

1. 870

/organism="Macaca mulatta"

/mol_type="mRNA"

/strain="Indian"

/db_xref="taxon:9544"

/clone="IBIUM:10227"

/sex="male"

/cell_type="mononuclear lymphocyte"

/dev_stage="adult"

/lab_host="E. coli SOLR"

/clone_lib="Katze MMSP"

/note="Organ: spleen; Vector: Uni-ZAP XR; Site 1: EcoR I;

Site 2: Xho I; Created from Stratagene ZAP-cDNA Synthesis

kit (catalog #200400) and ZAP-cDNA Gigapack III Gold

Cloning Kit (Catalog #200450)"

ORIGIN

Query Match 67.7%; Score 430.2; DB 7; Length 870;
Best Local Similarity 83.8%; Pred. No. 2.5e-113;
Matches 527; Conservative 0; Mismatches 88; Indels 14; Gaps 3;
QY 21 TGTAAAGTTTGTCTCTTCCAGAAATCAAAACTGTAAACTACAAATTTGTTCTTCAAATGGA 80
Db TTTTAAATTTGTTCTTCCAGAAATCGAAGTGTAAACTACAAATGGTCTTCAAATGGA 225
QY 81 GCACCAAGATGGAGTCCATGACTAAGATCCACCGTGGACCCCTGGACCGGCTCTAGCCC 140
Db GCCTCCGATGCGATCCATGACTAAGATCTACTGCGGACCCCTGGACCGGCTCTAATCC 285
QY 141 ATGCTCCGATGTTAATGACATTGAAGGACCCCTCCCGAGGAATCTCACTGCACACCC 200
Db ATGCTCTGATGTTGGTGACATCGAAGGACCCCTCCCGAGGAATCTCACTGCATGACC 345
QY 201 CCTACTATCCCAATTTCAAGCGGAGAGTGTAGAGCGGTCTATCAGCAACCTCCCAAC 260
Db CCTACTACGCTCAGTTTCAAGCGGAGAGTGTAGAGCGGTCTATCAGCAACCTCCCAAC 405
QY 261 AGCATTGGGTTTTCTCTTTGAGAGGGGGAGTGTAGAGACAGGACTAGCTGGATTCTTA 320
Db AGCATTGGGTTTTCTCTTTGAGAGGGGGAGTGTAGAGACAGGACTAGCTGGATTCTTA 465
QY 321 GGCCAAAGAAATCCCTAAGCCTAGCTGGGAAGTGTAGCTGCATCACCCTCTAAACATGG 380
Db GGCCAAAGAAATCCCTAAGCCTAGCTGGGAAGTGTAGCTGCATCACCCTCTAAACATGG 525
QY 381 GGCTTCAACTTAGCTTCACACCCGCAATC-----AGAGAGCTCACTAAATGCT 431
Db GGCTTCAACTTAGCTTCACACCCGCAATC-----AGAGAGCTCACTAAATGCT 585
QY 432 AATTAGGCAAAATAGGAGGTAAAGAAATAGCCAAATCATC--TATTGCTCTAGAGCACAG 489
Db AATTAGGCAAAATAGGAGGTAAAGAAATAGTCAATCACCCTTATCGCTCTAGAGCACAG 645
QY 490 CGGAGGGAAGAGGATTCGGGATATAAACCCAGGCAATTCGAGCGGGCAACGGCAACCCCT 549
Db GAGGAAGGAACAATGATAGGATATAAACCCAGGCAATTCGAGCGGGCAACGGCAACCCCT 705
QY 550 TT---GGGTCCCTCCCTTTGATGGCGCTCTGTTTTCACCTCTATTCTCTATTATAA 606
Db TTTTGTGTCCTCCCTCCCTTTTATGGAGCTCTGGGTTTCTCTCTATTATAATCTTGAAT 765


```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 440)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: AFCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
Plate: LCM752 row: j column: 01
High quality sequence stop: 440.
Location/Qualifiers
FEATURES
source
1..440
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGS:325728"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 21"
/notes="Organ: placenta; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGCAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

ORIGIN

```

Query Match 62.6%; Score 397.4; DB 2; Length 440;
Best Local Similarity 94.1%; Pred. No. 7.1e-104;
Matches 413; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 179 AGGAAATCTCACTGCACACCCCTACTATGCCCCCAATTCAGCGGGAGCAGTTAGAGCG 238
DB 1 AGGAAATCTCACTGCACACCCCTACTATGCCCCCAATTCAGCGGGAGCAGTTAGAGCG 60

QY 239 CTCATCAGCCAACTCCCAACAGCAGCTTGGGTTTCTGTGTGAGAGGGGGAGCTGAGAG 298
DB 61 GTGCTGGCCAACTCCCAACAGCAGCTTGGGTTTCTGTGTGAGAGGGGGAGCTGAGAG 120

QY 299 ACAGGACTAGCTGGATTCTTCAGGCCAACAGGAATCCCTTAAGCTAGCTGGGAGGTGA 358
DB 121 ACAGGACTAGCTGGATTCTTCAGGCCAACAGGAATCCCTTAAGCTAGCTGGGAGGTGA 180

QY 359 CTGATCCACCTTAAACATGGGCTTGCACCTTAGCTCACCACCGACCAATCAGAGAGC 418
DB 181 CCACATCCACCTTAAACACGGGGCTTGCACCTTAGCTCACCACCGACCAATCAGAGAGC 240

QY 419 TCACATAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCC 478
DB 241 TCACATAAATGCTAATTAGGCAAAATAGGAGGTAAAGAAATAGCAATCATCTATTGCC 300

QY 479 TGAGAGCAGCGGAGGAGCAGAGATCGGATATTAACCCAGGATTCGAGCGGCAAC 538
DB 301 TGAGAGCAGCGGAGGAGCAGAGATCGGATATTAACCCAGGATTCGAGCGGCAAC 360

QY 539 GGCACACCCCTTTGGGTCCTCCCTTTGTATGGCGCTCTGTTTCACTCTATTTCAC 598
DB 361 GGCACACCCCTTTGGGTCCTCCCTTTGTATGGAGCTCTGTTTCACTCTATTTCAC 420

QY 599 CTATTAAATCTTGCAACTG 617
DB 421 CTATTAAATCTTGCAACTG 439
```

RESULT 22

```

BI963185/c
LOCUS ie57c10.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
DEFINITION CDNA clone IMAGE:5670930 3', mRNA sequence.
ACCESSION BI963185
VERSION BI963185.1 GI:16337590
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 543)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Scarce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Plistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
Cardenas,M., Gibbons,M., McCann,R., Cole,R., Teagareishvili,R.,
Williams,T., Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brownjfas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 451.
Location/Qualifiers
FEATURES
source
1..543
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5670930"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/notes="Organ: Pancreas; Vector: pSPORII; Site 1: Not 1;
Site 2: Sal 1; Starting library constructed using
SuperScript plasmid library kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."
```

ORIGIN

```

Query Match 62.4%; Score 396; DB 4; Length 543;
Best Local Similarity 88.3%; Pred. No. 1.9e-103;
Matches 467; Conservative 0; Mismatches 51; Indels 11; Gaps 3;

QY 74 AAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGCACCCCTGGACCGGCTG 133
DB 543 AAATGGAGCACCAGATGGAGTCCATGACTAAGATCCACCGTGCACCCCTGGACCGGCTG 484

QY 134 CTAGCCCATGCTCGATGTTAATGACATTCGAGGCCCTCCCGAGGAATCTCAACTG 193
DB 483 CTAGCCCA-GCTCCACATTAATGACATCAAGGACCCCTCCGAGGAATCTCAACTG 425
```

Qy	194	CACAAACCCCTACTATGCCCCCAATTCAGCGGGAAGCAGTTAGAGCGGTCAATCAGGCCAACCT	253
Db	424	CATGACACCTACTACGCCCCCAATTCAGCAGGAAGCAGTTAGAGTGGTCGTTGGGCCAACCT	365
Qy	254	CCCCAAAGCAGCATTGGGTTTTCCTGTTGAGAGGGGGGNCCTAGAGACAGAGCTAGCTGGA	313
Db	364	CCCCAAAGCAGCATTGGGTTTTCCTGTTGAGAGGGGGGNCCTAGAGACAGAGCTAGCTGGA	306
Qy	314	TTTCTTAGGCCAACGAAGATCCCTAGCCCTAGCTGGGAAGGTGACTGCATCCACCTCTA	373
Db	305	TTTCTTAGGCCAACGAAGATCCCTAGCCCTAGCTGGGAAGGTGACTGCATCCACCTCTA	246
Qy	374	AACATGGGGCTTGCACCTTAGCTCACACCCGACCAAT-----CAGAGAGCTCACTA	424
Db	245	ACCCTAGGCTTGCACCTTAGCTCACACCCGACCAAT-----CAGAGAGCTCACTA	186
Qy	425	AAATGCTAATTAGGCAAAATAGAGAGTAAGAAATAGCCAAATCATCTATTTCCTGAGAG	484
Db	185	AAATGCTAATTAGGCAAAATAGAGAGTAAGAAATAGCCAAATCATCTATTTCCTGAGAG	126
Qy	485	CACAGCGGAGGGAACAGGATCGGGATATAAACCCAGGCATTCAGAGCGGCAACGGCAAC	544
Db	125	CACAGCGGAGGGAACAGGATCGGGATATAAACCCAGGCATTCAGAGCGGCAACGGCAAC	66
Qy	545	CCCCTTTGGGTCCTCCCTCTGTTATGGCGCTCTGTTTTCACCTCTATT	593
Db	65	CCACTTTGGGTCCTCCCTCTGTTATGGCGCTCTGTTTTCACCTCTATT	17
RESULT 23			
LOCUS	176		
DEFINITION	176	966 bp mRNA linear EST 28-APR-2004	
ACCESSION	176	966 bp mRNA linear EST 28-APR-2004	
VERSION	176	966 bp mRNA linear EST 28-APR-2004	
KEYWORDS	176	966 bp mRNA linear EST 28-APR-2004	
SOURCE	176	966 bp mRNA linear EST 28-APR-2004	
ORGANISM	176	966 bp mRNA linear EST 28-APR-2004	
REFERENCE	176	966 bp mRNA linear EST 28-APR-2004	
AUTHORS	176	966 bp mRNA linear EST 28-APR-2004	
TITLE	176	966 bp mRNA linear EST 28-APR-2004	
COMMENT	176	966 bp mRNA linear EST 28-APR-2004	

Query Match	62.2%	Score 395.2	DB 5	Length 966
Best local Similarity	79.6%	Pred. No. 3.7e-103		
Matches 465	Conservative 22	Mismatches 84	Indels 13	Gaps 2
QY	1	CCCTGATCTTTAAACCTCTTGTAAAGTTTGTCTCTTCAGAAATCAAAACCTGATAAACTA	60	
Db	574	CCCTGATCTTTAAACCTCTTGTAAAGTTTGTCTCTTCAGAAATCAAAACCTGATAAACTA	515	
QY	61	CAAAATTGTTCTTCAAAATGGAGCAGCAGATGGAGTCCATGACTTAAGATCCACCGTGGACCC	120	
Db	514	-----CAAAATGGAGCCAGATGCACTCAGACTAGGDTCTACCGCGGCCCC	467	
QY	121	CTGGACGGGCTGCTAGCCCATGCTCGATGTTTAATGACATTAAGAGCACCCCTCCCGAG	180	
Db	466	CTGGACGGGCTGCTGSGCCCCGTTGATGTTTAATGACATTAAGAGCACCCCTCCCTGGG	407	
QY	181	GAATCTCAACTGCACAAACCCCTACTATGCCCCAAATTCAGCGGAGACGATTAGACGGT	240	
Db	406	GCAATCTCAGCTGCACMCCCTCTACTACGCCCAATTCGCGAGACGATTTRGGCGGT	347	
QY	241	CATCAGCAACCTCCCAACAGCACACTTGGGTTTTCCTGTTGAGAGGGGGGACTTGAGAGAC	300	
Db	346	CGTCGCCCAACCTCCCGACRGCACTTGGTTTTCTGTTGAGATGGGGGACTTGAGAGAC	287	
QY	301	AGACTAGCTGGAATTTCTTAGGCCAACGAAGAAATCCCTAAGCTAGCTGGAAAGGTGACT	360	
Db	286	MGACTGGCTGGATTTCTTAGCTGCACTTAAGAAATCCTARGCCTRGCTGGAAAGGTGACC	227	
QY	361	GCATCACCTCTAAACATGGGGCTTGCATTTAGCTTACACCCGACCAATCAGAGAGCTC	420	
Db	226	ACATCACCTTTAAACMCGGGGCTTCCCACTTSGCTCACCCCTGCCCAATCRGGGGGCTC	167	
QY	421	ACTAAATGCTAATTAGGCCAAAAATAGGAGGTAAAGAAATAGC-CAATCATCTATTGCGCT	479	
Db	166	ACTAAATGCTAMTYCGGCCCGSCSSSSVSSSSCCCTCGCGMSYCWTTATTGCGCT	107	
QY	480	GAGAGCACAGCGGGAGGACAGGATCGGATATAAACCCAGGCATTCGAGCGCGCAACG	539	
Db	106	GAGAGCACAGCMGGAGGACAAATGATCGGATATAAACCCAMGTCTTCGCGCGCGGCCG	47	
QY	540	GCAACCCCTTTGGGTCCCTTCTGTTGATGGCGGCTCTGTT	583	
Db	46	GCCCCCCCCCTTTGGGTCCCTTCTGTTGATGGCGGCTCTGTT	3	
RESULT 24				
AG058970/c				
LOCUS	AG058970	651 bp	DNA	linear
DEFINITION	Pan troglodytes DNA, clone: PTB-046A08.R, genomic survey sequence.			
ACCESSION	AG058970			
VERSION	AG058970.1 GI:16596431			
KEYWORDS	GSS.			
SOURCE	Pan troglodytes (chimpanzee)			
ORGANISM	Pan troglodytes			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.			
AUTHORS	1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.			
TITLE	BAC end sequences of Library PTB			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 651)			
TITLE	Totoki, Y., Watanabe, H. and Sakaki, Y.			
JOURNAL	Direct Submission			
REFERENCE	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan			
AUTHORS	[E-mail:chimps@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170]			
TITLE	Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.			
COMMENT				

405	ACCAATCA-----GAGAGCTCACTAAATGCTAAATTAGGCGAAAAATAGGAGGTAAA	455
247	ACCAATCAGATAGTAGGAGAGAGCTCACTAAATGCTAAATTAGGCGAACCAACAGGAGGTAAA	188
456	GAATAGCCCAATCATCTATTGCTGAGAGCAGCAGCGGAGGAGGACNAGGATCGGATATAA	515
187	GAATAGCCCAATCATCTGCTACCTGAAAGCAGTGGGAGGAGACAAATGATTGGGATATAA	128
516	ACCCAGGCATTGAGCGCGCAACGGCAACCCCTTTGGGTCCCTCCCTCTTTGTATGGCGC	575
127	ACCCAGGCATTGAGCCAGCAATGCAACCCCTTTGGGTCCCTCTTTGTATGGGAG	68
576	CTCTG 580	
67	CTCG 63	
RESULT 27		
AA837267/c		
LOCUS		
DEFINITION	443 bp mRNA linear EST 31-MAR-1998	
	od26b10.s1 NCI CGAP GCBI Homo sapiens cDNA clone IMAGE:1369051	
	similar to contains PTR7.tl PTR7 repetitive element ; mRNA	
	sequence.	
ACCESSION	AA837267	
VERSION	AA837267.1	
KEYWORDS	GI:2912466	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 443)	
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),	
JOURNAL	Tumor Gene Index	
COMMENT	Unpublished (1997)	
	Contact: Robert Strausberg, Ph.D.	
	Email: cgapbs-remail.nih.gov	
	Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,	
	Ph.D., Gerald Marti, M.D.	
	CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima	
	Bonaldo, Ph.D.	
	CDNA Library Arrayed by: Greg Lennon, Ph.D.	
	DNA Sequencing by: Washington University Genome Sequencing Center	
	Clone distribution: NCI-CGAP clone distribution information can be	
	found through the I.M.A.G.E. Consortium/LLNL at:	
	www.bio.llnl.gov/bbrp/image/image.html	
	Insert Length: 1375 Std Error: 0.00	
	Seq primer: -40m13 fwd. ET from Amersham.	
FEATURES		
source	1. 443	
	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:1369051"	
	/tissue_type="germinal center B cell"	
	/lab_host="DH10B"	
	/clone_lib="NCI CGAP GCBI"	
	/note="Vector: pTR73D-Pac (Pharmacia) with a modified	
	polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand	
	was prepared from human tonsillar cells enriched for	
	germinal center B cells by flow sorting (CD20+, IgD-),	
	provided by Dr. Louis M. Staudt (NCI). Dr. David Allman	
	(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was	
	primed with a Not I - oligo(dT) primer	
	15'-GTGTACCAATCTGAATGGGCGCGCTCATTTTTTTTTTTT-3'.	
	1. Double-stranded cDNA was ligated to Eco RI adaptors	
	(Pharmacia), digested with Not I and cloned into the Not I	
	and Eco RI sites of the modified pTR73 vector. Library	
	went through one round of normalization, and was	
	constructed by Bento Soares and M. Fatima Bonaldo."	
ORIGIN		
405	ACCAATCA-----GAGAGCTCACTAAATGCTAAATTAGGCGAAAAATAGGAGGTAAA	455
247	ACCAATCAGATAGTAGGAGAGAGCTCACTAAATGCTAAATTAGGCGAACCAACAGGAGGTAAA	188
456	GAATAGCCCAATCATCTATTGCTGAGAGCAGCAGCGGAGGAGGACNAGGATCGGATATAA	515
187	GAATAGCCCAATCATCTGCTACCTGAAAGCAGTGGGAGGAGACAAATGATTGGGATATAA	128
516	ACCCAGGCATTGAGCGCGCAACGGCAACCCCTTTGGGTCCCTCCCTCTTTGTATGGCGC	575
127	ACCCAGGCATTGAGCCAGCAATGCAACCCCTTTGGGTCCCTCTTTGTATGGGAG	68
576	CTCTG 580	
67	CTCG 63	
RESULT 27		
AA837267/c		
LOCUS		
DEFINITION	443 bp mRNA linear EST 31-MAR-1998	
	od26b10.s1 NCI CGAP GCBI Homo sapiens cDNA clone IMAGE:1369051	
	similar to contains PTR7.tl PTR7 repetitive element ; mRNA	
	sequence.	
ACCESSION	AA837267	
VERSION	AA837267.1	
KEYWORDS	GI:2912466	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 443)	
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),	
JOURNAL	Tumor Gene Index	
COMMENT	Unpublished (1997)	
	Contact: Robert Strausberg, Ph.D.	
	Email: cgapbs-remail.nih.gov	
	Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,	
	Ph.D., Gerald Marti, M.D.	
	CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima	
	Bonaldo, Ph.D.	
	CDNA Library Arrayed by: Greg Lennon, Ph.D.	
	DNA Sequencing by: Washington University Genome Sequencing Center	
	Clone distribution: NCI-CGAP clone distribution information can be	
	found through the I.M.A.G.E. Consortium/LLNL at:	
	www.bio.llnl.gov/bbrp/image/image.html	
	Insert Length: 1375 Std Error: 0.00	
	Seq primer: -40m13 fwd. ET from Amersham.	
FEATURES		
source	1. 443	
	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:1369051"	
	/tissue_type="germinal center B cell"	
	/lab_host="DH10B"	
	/clone_lib="NCI CGAP GCBI"	
	/note="Vector: pTR73D-Pac (Pharmacia) with a modified	
	polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand	
	was prepared from human tonsillar cells enriched for	
	germinal center B cells by flow sorting (CD20+, IgD-),	
	provided by Dr. Louis M. Staudt (NCI). Dr. David Allman	
	(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was	
	primed with a Not I - oligo(dT) primer	
	15'-GTGTACCAATCTGAATGGGCGCGCTCATTTTTTTTTTTT-3'.	
	1. Double-stranded cDNA was ligated to Eco RI adaptors	
	(Pharmacia), digested with Not I and cloned into the Not I	
	and Eco RI sites of the modified pTR73 vector. Library	
	went through one round of normalization, and was	
	constructed by Bento Soares and M. Fatima Bonaldo."	
ORIGIN		
405		

Query Match 61.8%; Score 392.6; DB 1; Length 443;
Best Local Similarity 94.1%; Pred. No. 1.8e-102;
Matches 418; Conservative 0; Mismatches 25; Indels 1; Gaps 1;
184 ATCTCACTGCAACACCCCTACTATATCCCAATTCAGCGGAGAGAGTAGAGCGGTCTAT 243
|||||
443 ATCTCACTGCAACACCCCTACTATATCCCAATTCAGCGGAGAGAGTAGAGCGGTCTAT 384
|||||
244 CAGCAGACCTCCCAACAGACACTTGGTTTCTCTTTGAGAGGGGAGCTGAGAGACAGG 303
|||||
383 CAGCAGACCTCCCAACAGACACTTGGTTTCTCTTTGAGAGGGGAGCTGAGAGACAGG 324
|||||
304 ACTAGCTGGATTTCTTAGGCAACAGAGATCCCTTAAGCCTAGCTGGAGAGGTGATGCA 363
|||||
323 ACTAGCTGGATTTCTTAGGCGGATTAAGATCCCTTAACCTAGCTGGAGAGGTGACCGCA 264
|||||
364 TCCACCTCTAAACATGGGGCTTCCAACTTAGCTTACACCCGACCAATCAGAGAGCTCACT 423
|||||
263 TCCACCTTTAAACAGGGGCTTGCACCTTAGCTTACACCCCAACCAATCAGAGAGCTCACT 204
|||||
424 AAATGCTAATAGGCAAAATAGGAGGTAAAGAAATAGCAATATGCTATTCCTCTGAGA 483
|||||
203 AAATGCTAATAGGCAAAATAGGAGGTAAAGAAATAGCAATATGCTATTCCTCTGAGA 144
|||||
484 GCACAGCGGAGGACAAAGATCGGATATAAAACCCAGGCAATTCGAGCGGCAACCGGCA 543
|||||
143 GCACAGTGGAGGACAAAGATTCGATATATAACCCAGGCAATTCGAGCGGCAACCGGCA 85
|||||
544 CCGCTTTGGGTCCCTCCCTTTGATAGGCGCTCTGTTTCTCTATTTCTCTATTTCTCTATT 603
84 CCGCTTTGGGTCCCTCCCTTTGATAGGCGCTCTGTTTCTCTATTTCTCTATTTCTCTATT 25
604 AAATCTTGCACCTGAAAAAATAA 627
24 AAATCTTGCACCTGAAAAAATAA 1

RESULT 28
A0892947/c
LOCUS
DEFINITION
HS 3131 B2 B04 MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3131 Col=8 Row=J, genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 712)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
10449764
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: <http://www.htsc.washington.edu>
Plate: 3131 row: J column: 8
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 712.

FEATURES
source

Location/Qualifiers
1..712
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="plate=3131 Col=8 Row=J"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

ORIGIN

Query Match 61.0%; Score 387.2; DB 8; Length 712;
Best Local Similarity 82.2%; Pred. No. 7.2e-101;
Matches 484; Conservative 0; Mismatches 94; Indels 11; Gaps 3;
QY 56 AACTACAATTCCTTCAATGGAGACACAGATGGAGTCCATGACTAAGATCCACCGTG 115
|||
Db 699 AAGTTACAAGGTTTCTNCAAGGATCCCAAGATCAGTCTAT-ACCTCAATCTACCCGG 641
|||
QY 116 GACCCCTGGACCGGCTGCTAGCCCATGCTCCGATGTTTAATGACATTTGAAGGACACCCCTC 175
|||
Db 640 ACCCTTGGACCGGCTGCTAGTCCATGCTTCGATGTTGATGATATCAAAAGCAGCCCTC 581
|||
QY 176 CCGAGGAATCTCACTGCAACACCCCTACTATGCCCAATTCAGCGGGAAGCAGTTAGA 235
|||
Db 580 CCGAGGAATCTCAAGTGCATGACCCCTTAGT-TGCACACAGTTCAGCAGGAAGAGTTAGA 522
|||
QY 236 GGGTCTATCAGCAACCTCCCAACAGCACTTGGGTTTTCTCTTTGAGAGGGGGAGCTGA 295
|||
Db 521 GGGCGGTTGGCCACCTCCCAATAGTACTTGGGTTTTCTCTTTGAGAGGGGTTGCTGA 462
|||
QY 296 GAGACAGGACTAGCTGGATTCTTAGGCCAACAGAGAAATCCCTAGCCCTAGCTGGGAAGG 355
|||
Db 461 GAGACAGGACTAGCTGGATTCTTAGGCCGACTAAGAAATCCCTAGCCCTAGCTGGGAAGG 402
|||
QY 356 TGACTGCTCCACCTCTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAAT----- 410
|||
Db 401 TGACTGCTCCACCTTTAAACACAGGGGCTTGCAACTTAGCTCACACCCGACCAATGAGGT 342
|||
QY 411 ----CAGAGAGGCTCACTAAATAGCTAAATAGGCAAAATAGGAGGTAAAGAAATAGCCAA 466
|||
Db 341 AGTAAAGAGGCTCACTAAATAGCTAAATAGGCAAAATAGGAGGTAAAGAAATAGCCAA 282
|||
QY 467 TCATCTATTGCTGAGAGCAGCAGCGGAGGACAGGATCGGATATTAACCCAGGCATT 526
|||
Db 281 TCATCTATCACTGAGAGCAGCAGCGGAGGACAAATGATCAGGATATAACCCAGGCATT 222
|||
QY 527 CGAGCGGCAACGGCAACCCCTTTGGGTCCTCTCTCTTTGATGGCGCTCTGTTTCA 586
|||
Db 221 CTAGCGGCAACGGCTACCTCTTTGGGTACCTCTCTCTTTGATGGAGCTCTGTTTCA 162
|||
QY 587 CTCTATTTCACTCTATTAAATCTTGCAACTGAAAAAATAAAAAA 635
|||
Db 161 CTCTATTAATCTTGCAAGACAAAAACCAACCAACCAACCAAA 113
|||

RESULT 29

A1128526/c

LOCUS

DEFINITION

A1128526

436 bp mRNA linear EST 27-OCT-1998

GC61h10.x1 Soares placenta 8to9weeks 2NbhP8to9W Homo sapiens CDNA

clone IMAGE:171417.3' similar to contains PTR7.b1 PTR5 repetitive

element ;, mRNA sequence.

A1128526

A1128526.1 GI:3597040

EST.

KEYWORDS

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 436)

NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL
COMMENT

Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 720 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 428.
Location/Qualifiers

FEATURES
source
1. .436
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1714147"
/dev stage="two placentae: one from 8 weeks and another
from 9 weeks post conception"
/lab host="DH10B (ampicillin resistant)"
/clone lib="Soares_placenta_8to9weeks_2Nbp8to9W"
/notes="Organ: placenta; Vector: p7T3D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTCAAGTCGAGCGCGCGCATTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7T3 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo."

ORIGIN

Query Match 60.9%; Score 386.8; DB 1; Length 436;
Best Local Similarity 93.7%; Pred. No. 8.4e-101;
Matches 403; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 206 TATGCCCAATTCAGCGGAGCAGTTAGAGCGGTCAATCAGCAACTCCCAACAGCAC 265
Db 436 TACGCCCAATTCAGCAGGAGCAGTTAGAGCGGTCTGCGCCAACTCCCAACAGCAC 377

QY 266 TTGGGTTTTCTGTGAGAGGGGGAGCTGAGAGACAGGACTAGCTGGATTTCTAGGCCA 325
Db 376 TTAGGTTTTCTGTGAGATGGGGAGCTGAAGACAGGAGTCTGGATTTCTAGGCTG 317

QY 326 ACAGAAATCCCTAAGCTAGCTGGGAAGTGACTGATCCACTCTAAACATGGGGCTT 385
Db 316 ACTAAGAATCCCTAAGCTAGCTGGGAAGTGACTGATCCACTCTAAACATGGGGCTT 257

QY 386 GCAACTTAGCTACACCGCAACATCAGAGAGCTCACTAAATGCTAAATTAGGCAAAAT 445
Db 256 GCAACTTAGCTACACCTGACCAATCAGAGAGCTCACTAAATGCTAAATTAGGCAAGAC 197

QY 446 AGAGGTAAGAAATAGCCAATCATCTATTGCTGAGAGACACAGCGGGAGGACAAGGAT 505
Db 196 AGAGGTAAGAAATAGCCAATCATCTATTGCTGAGAGACACAGCAGGAGGACATGAT 137

QY 506 CGGGATATAAACCCAGGCATTCAGCGGCAACCGCAACCCCTTTGGGTCCCTCCCTT 565
Db 136 CGGGATATAAACCCAAAGTTTCAGAGCGGCAACCGCAACCCCTTTGGGTCCCTCCCTT 77

QY 566 TGTATGGGCGCTGTGTTTCACTCTATTTCACCTCTATTAAATCTTCAACTGAAAAAAA 625
Db 76 TGTATGGGAGCTGTGTTTTCATCTATTTCACCTCTATTAAATCTTCAACTGCAAAAAA 17

QY 626 AAAAAAAA 635
Db 16 AAAAAAAA 7

RESULT 30
AI393478/c
LOCUS
DEFINITION
IMAGE:2111766 3', mRNA sequence.
AI393478
ACCESSION

AI393478.1 GI:4223025
EST.
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 446)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1061 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 445.
Location/Qualifiers

FEATURES
source
1. .446
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2111766"
/lab host="DH10B"
/clone lib="Soares_NFL_T_GBC_S1"
/notes="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 60.7%; Score 385.2; DB 1; Length 446;
Best Local Similarity 92.5%; Pred. No. 2.5e-100;
Matches 419; Conservative 0; Mismatches 24; Indels 10; Gaps 1;

QY 183 AATCTCAACTGCACAACTTATGCCCCCAATTCAGCGGAGCAGTTAGAGCGTCA 242
Db 446 AATTTCAACTGTGCAACCTTANTATGCCCAATTCAGCAGGAGCAGTTACAGCGTCA 387

QY 243 TCAGCCAACTCCCAACAGCAGCTTGGGTTTTCTGTTGAGAGGGGGAGCTGAGAGACAG 302
Db 386 TCAGCCAACTCCCAACAGCAGCTTGGGTTTTCTGTTGAGAGGGGGAGCTGAGAGACAA 327

QY 303 GACTAGCTGGATTTCTTAGGCCAAACGAGAGTCCCTAAGCCTAGCTGGGAAGTGTGCTGC 362
Db 326 GACTAGCTGGATTTCTTAGGCCAACTAAGATCCCTAAGCCTAGCTGGGAAGTGTGCTGC 267

QY 363 ATCCACTCTAAACATGGGGCTTGCACCTTAGCTCACACCCGACCAATCAGAGAGTCTAC 422
Db 266 ATCCACTTTAAACATGGGGCTTGCACCTTAGCTCACACCCGACCAATCAGAGAGTCTAC 207

QY 423 TAAATGCTAATTAGGCAAAAAATAGAGGTAAGAAATAGCCAAATCATCTATTGCGCTGAG 482
Db 206 TAAATGCTAATTAGGCAAAAAATAGAGGTAAGAAATAGCCAAATCACCTTTTGGCTGAG 147

QY 483 AGCAGAGCGGAGGAGCAAGGATCGGGATATAACCCAGGCAATTCAGGCGCGCAAGGCA 542
Db 146 AGCAGAGCGGAGGAGCAAGGATCGGGATATAACCCGCGCATTAAGCCGCGCAATGGCA 87

QY 543 ACCCCCTTTGGGTCCCTCCCTTTGATGGGGCTCTGTTTTTCACTCTATTTCACCTCTAT 602
Db 86 ACCCCCTTTGGGTCCCTCCCTTTGATGGAGCTCTGT-----TTTCACTCTAT 37

QY 603 TAAATCTGCAACTGAAAAAAGAAAAA 635

Db 36 TAAATCTTGCAACTGCAAAAAAAAAAAAAAAAAAAAA 4

Search completed: January 22, 2005, 19:31:32
Job time : 1902.54 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2005, 10:09:14 ; Search time 6212.46 Seconds
(without alignments)
11273.477 Million cell updates/sec

Title: US-09-319-156B-9
Perfect score: 1481
Sequence: 1 atggccctccctatcatac.....gtctacaatgggaacccca 1481

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hgt: *
3: gb_in: *
4: gb_cm: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1481	100.0	1481	6 AR344387	AR344387 Sequence
2	1481	100.0	1481	6 AX001027	AX001027 Sequence
3	1481	100.0	1481	6 BD136197	BD136197 Retrovira
4	1481	100.0	1932	14 AF127228	AF127228 Multiple
5	1468.2	99.1	162579	9 AL390039	AL390039 Human DNA
6	1466.6	99.0	2055	6 CQ719381	CQ719381 Sequence
7	1461.8	98.7	1629	6 BD252040	BD252040 Method fo
8	1461.8	98.7	1629	6 AX036992	AX036992 Sequence
9	1461.8	98.7	1629	14 AF331500	AF331500 Multiple
10	1461.8	98.7	172918	2 AC023366	AC023366 Homo sapi
11	1389.8	93.8	2074	6 AX478550	AX478550 Sequence
12	1341.8	90.6	10230	9 AY101589	AY101589 Gorilla g
13	1341.8	90.6	184675	9 AC145964	AC145964 Pan trogl
14	1340.2	90.5	10229	9 AY101586	AY101586 Pan trogl
15	1340.2	90.5	10229	9 AY101587	AY101587 Pan trogl
16	1340.2	90.5	10230	9 AY101588	AY101588 Gorilla g
17	1338.6	90.4	10246	9 AY101593	AY101593 Hylobates
18	1338.6	90.4	10248	9 AY101592	AY101592 Hylobates
19	1335.4	90.2	2694	9 AF520478	AF520478 Homo sapi

20	1335.4	90.2	2694	9 AF520487S2	AF520488 Homo sapi
21	1335.4	90.2	2694	9 AF520489S2	AF520490 Homo sapi
22	1335.4	90.2	2694	9 AF520505S2	AF520506 Homo sapi
23	1335.4	90.2	2694	9 AF520517S2	AF520518 Homo sapi
24	1335.4	90.2	2694	9 AF520519S2	AF520520 Homo sapi
25	1335.4	90.2	2694	9 AF520521S2	AF520522 Homo sapi
26	1335.4	90.2	2694	9 AF520529S2	AF520530 Homo sapi
27	1335.4	90.2	2694	9 AF520561S2	AF520562 Homo sapi
28	1335.4	90.2	2930	6 AX355872	AX355872 Sequence
29	1335.4	90.2	2930	6 AF208161	AF208161 Homo sapi
30	1335.4	90.2	2946	6 AR177269	AR177269 Sequence
31	1335.4	90.2	2946	6 BD232464	BD232464 Secreted
32	1335.4	90.2	2946	6 BD086133	BD086133 Secreted
33	1333.8	90.1	1624	9 AF506835	AF506835 Homo sapi
34	1333.8	90.1	1860	9 AF511360	AF511360 Homo sapi
35	1333.8	90.1	2055	6 BD221827	BD221827 Nucleic s
36	1333.8	90.1	2055	6 AX007999	AX007999 Sequence
37	1333.8	90.1	2599	6 BD221806	BD221806 Nucleic s
38	1333.8	90.1	2599	6 AX007978	AX007978 Sequence
39	1333.8	90.1	2694	9 AF520481S2	AF520482 Homo sapi
40	1333.8	90.1	2694	9 AF520483S2	AF520484 Homo sapi
41	1333.8	90.1	2694	9 AF520485S2	AF520486 Homo sapi
42	1333.8	90.1	2694	9 AF520491S2	AF520492 Homo sapi
43	1333.8	90.1	2694	9 AF520493S2	AF520494 Homo sapi
44	1333.8	90.1	2694	9 AF520495S2	AF520496 Homo sapi
45	1333.8	90.1	2694	9 AF520497S2	AF520498 Homo sapi

ALIGNMENTS

RESULT 1
AR344387
LOCUS AR344387 1481 bp mRNA linear PAT 17-AUG-2003
DEFINITION Sequence 105 from patent US 6582703.
ACCESSION AR344387
VERSION AR344387.1 GI:33740328
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1481)
AUTHORS Perron,H., Beseme,F., Bedin,F., Paranhos-Baccala,G., Komurian-Pradel,F., Jollivet-Reynaud,C. and Mandrand,B.
TITLE Isolated nucleotide sequences associated with multiple sclerosis or rheumatoid arthritis and a process of detecting
JOURNAL Patent: US 6582703-A 105 24-JUN-2003;
FEATURES
source Location/Qualifiers
1..1481
/organism="unknown"
/mol_type="mRNA"

ORIGIN	Query Match	Best Local Similarity	Matches 1481;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	ATGGCCCTCCCTTATCATCTTTCTTTACTGTTCTCTTACCCCTTCGCTCTCACT	60								
Db	1	ATGGCCCTCCCTTATCATCTTTCTTTACTGTTCTCTTACCCCTTCGCTCTCACT	60								
Qy	61	GCACCCCTCCATGCTGTGTAACAACAGTAGTCCCTTACCAAGAGTTTCTATGAAGA	120								
Db	61	GCACCCCTCCATGCTGTGTAACAACAGTAGTCCCTTACCAAGAGTTTCTATGAAGA	120								
Qy	121	ACGCGCTTCTCGGAATATTGATGCCCATCATATAGGAGTTTATAGGAACTCC	180								
Db	121	ACGCGCTTCTCGGAATATTGATGCCCATCATATAGGAGTTTATAGGAACTCC	180								
Qy	181	ACCTTCACGTCCACACCCATATGCCCGCACTGTATAACTCTGCACTCTTTGCATG	240								
Db	181	ACCTTCACGTCCACACCCATATGCCCGCACTGTATAACTCTGCACTCTTTGCATG	240								


```
Db 481 CATACTCGCTGGTGACCTATTATATACACCCCTCACTCGCTCCATGAGGTCTAGCC 540
Qy 541 CAAAACCCCTACTAATCTGTGGATGTGCTCCCTGCACCTTCAGGCCCATACATTTCATC 600
Db 541 CAAAACCCCTACTAATCTGTGGATGTGCTCCCTGCACCTTCAGGCCCATACATTTCATC 600
Qy 601 CCTGTTCTCTGAACCAATGGAACAATTCAGCACAGAAATAAACACCCTCCGTTTGTAGTA 660
Db 601 CCTGTTCTCTGAACCAATGGAACAATTCAGCACAGAAATAAACACCCTCCGTTTGTAGTA 660
Qy 661 GGACCTCTGTTTCCAAATCGGAATAAACCAATACCTCAAACTCAGCTGTGTAAATTT 720
Db 661 GGACCTCTGTTTCCAAATCGGAATAAACCAATACCTCAAACTCAGCTGTGTAAATTT 720
Qy 721 AGCAATCTATAGACACAACCAAGCTCCCAATGATCATCAGTGGGTAAACACCCTCCACACGA 780
Db 721 AGCAATCTATAGACACAACCAAGCTCCCAATGATCATCAGTGGGTAAACACCCTCCACACGA 780
Qy 781 ATAGTCTGCTACCTCAGCAATATTTTGTCTGTGGTACCTCAGCCTATCATTTGTTG 840
Db 781 ATAGTCTGCTACCTCAGCAATATTTTGTCTGTGGTACCTCAGCCTATCATTTGTTG 840
Qy 841 AATGGCTCTTCAAGATCTATGTGCTCTCTCTCATTTCTAGTGCCCTATGACCATCTAC 900
Db 841 AATGGCTCTTCAAGATCTATGTGCTCTCTCTCATTTCTAGTGCCCTATGACCATCTAC 900
Qy 901 ACTGAACAAGATTATACAAATCATGTCGTACTCAAGCCCAACCAACCAAGAGTACCCATT 960
Db 901 ACTGAACAAGATTATACAAATCATGTCGTACTCAAGCCCAACCAACCAAGAGTACCCATT 960
Qy 961 CTTCTCTTTTATCAGACAGAGTCTAGGACAGACTAGGTACTGGCATTTGGCAGTATC 1020
Db 961 CTTCTCTTTTATCAGACAGAGTCTAGGACAGACTAGGTACTGGCATTTGGCAGTATC 1020
Qy 1021 ACAACCTCTACTCAGTTCTACTCAAACTATCTCAAGAAATAATGTGTGACATGGACAG 1080
Db 1021 ACAACCTCTACTCAGTTCTACTCAAACTATCTCAAGAAATAATGTGTGACATGGACAG 1080
Qy 1081 GTCACTGACTCCCTGGTCACTTGCAGATCAACTTAACCTCCCTAGCAGCAGTAGTCCTT 1140
Db 1081 GTCACTGACTCCCTGGTCACTTGCAGATCAACTTAACCTCCCTAGCAGCAGTAGTCCTT 1140
Qy 1141 CAAAATCGAAGAGCTTTAGACTTGTCTAAACCGCAAAAGAGGGGGAACCTGTTATTTTA 1200
Db 1141 CAAAATCGAAGAGCTTTAGACTTGTCTAAACCGCAAAAGAGGGGGAACCTGTTATTTTA 1200
Qy 1201 GGAGAGAACGCTGTTATTATGTTTAACTCAATCCAGATTTCTCACTGAGAAAGTTAAAGAA 1260
Db 1201 GGAGAGAACGCTGTTATTATGTTTAACTCAATCCAGATTTCTCACTGAGAAAGTTAAAGAA 1260
Qy 1261 ATTTCGAGATCGAATACAAATGTAGACGAGAGCTTCAAAACACCGAACGCTGGGGCTC 1320
Db 1261 ATTTCGAGATCGAATACAAATGTAGACGAGAGCTTCAAAACACCGAACGCTGGGGCTC 1320
Qy 1321 CTAGCCAAATGATGCTGGTCTCCCTTCTTAGGACCTCTAGCAGCTCTAAATTTG 1380
Db 1321 CTAGCCAAATGATGCTGGTCTCCCTTCTTAGGACCTCTAGCAGCTCTAAATTTG 1380
Qy 1381 TTAATCTCTTGTGACCCCTGATCTTAACTCTCTGTTAAAGTTGCTCTCTCCAGAAAT 1440
Db 1381 TTAATCTCTTGTGACCCCTGATCTTAACTCTCTGTTAAAGTTGCTCTCTCCAGAAAT 1440
Qy 1441 GAAGCTGTAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
Db 1441 GAAGCTGTAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
```

RESULT 3
BD136197
LOCUS
DEFINITION
BD136197 1481 bp DNA linear PAT 18-SEP-2002
Retroviral nucleic material and nucleotide fragments, in
particular, associated with multiple sclerosis and/or rheumatoid

arthritis, for diagnostic, prophylactic and therapeutic uses.
BD136197
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
ORIGIN
Query Match 100.0%; Score 1481; DB 6; Length 1481;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGCGCCCTCCCTTATCATCTTTTCTCTTTTACTGTTCTTACCCCTTTCCCTTTTCGCTCTCCT 60
Db 1 ATGCGCCCTCCCTTATCATCTTTTCTCTTTTACTGTTCTTACCCCTTTCCCTTTTCGCTCTCCT 60
Qy 61 GCACCCCTCCATGCTGTCAACACGAGTAGCTCCCTTTACCAAGAGTTTCTTATGAAGA 120
Db 61 GCACCCCTCCATGCTGTCAACACGAGTAGCTCCCTTTACCAAGAGTTTCTTATGAAGA 120
Qy 121 ACSCGGTCTCTGGAATAATTTGATGCCCATATATAGGAGTTTATTAAGGGAATCC 180
Db 121 ACSCGGTCTCTGGAATAATTTGATGCCCATATATAGGAGTTTATTAAGGGAATCC 180
Qy 181 ACCTTCACCTGCCACACCCATATGCCCGCACTGCTATAACTCTGCCACTCTTTGCATG 240
Db 181 ACCTTCACCTGCCACACCCATATGCCCGCACTGCTATAACTCTGCCACTCTTTGCATG 240
Qy 241 CATGCAAAATCTCATTTATTGGACAGGGAATAATGATTAACTTAGTTGCTCTGGAGGACTT 300
Db 241 CATGCAAAATCTCATTTATTGGACAGGGAATAATGATTAACTTAGTTGCTCTGGAGGACTT 300
Qy 301 GGAGCCACTGTCTGTTGGACTTACTTCAACCCATACCCAGTATGCTGATGGGGTGGAAAT 360
Db 301 GGAGCCACTGTCTGTTGGACTTACTTCAACCCATACCCAGTATGCTGATGGGGTGGAAAT 360
Qy 361 CAGGTCAGGCAGAGAAAACAGTAAGGAGCAATCTCCCAACTGACCCCGGAGCAT 420

Qy	361	CAAGGTCAGGCAAGAGAAAAACAAGATAAAGGAACAATCTCCCAACTGACCCCGGGACAT	420
Db	812		
		CAAGGTCAGGCAAGAGAAAAACAAGATAAAGGAACAATCTCCCAACTGACCCCGGGACAT	871
Qy	421	AGCACCCCTAGCCCTCAAAAGGACTAGTTCTCTCAAAACTACATGATGAACCCCTCCGTACC	480
Db	872		
		AGCACCCCTAGCCCTCAAAAGGACTAGTTCTCTCAAAACTACATGATGAACCCCTCCGTACC	931
Qy	481	CATACCTCGCCTGTGTGAGCCTATTTAATACCAACCCCTCACTCGGCTCCATGAGGTCACGCC	540
Db	932		
		CATACCTCGCCTGTGTGAGCCTATTTAATACCAACCCCTCACTCGGCTCCATGAGGTCACGCC	991
Qy	541	CAAAACCTTACTAATCTGTTGATGTGCTCCCTCGCACTTCAGGCCATATACATTTCAATC	600
Db	992		
		CAAAACCTTACTAATCTGTTGATGTGCTCCCTCGCACTTCAGGCCATATACATTTCAATC	1051
Qy	601	CCTGTCCTCGAAACAAATGGAAACAACTTCAGCACAGAAATAAACACCACTTCGCTTTTAGTA	660
Db	1052		
		CCTGTCCTCGAAACAAATGGAAACAACTTCAGCACAGAAATAAACACCACTTCGCTTTTAGTA	1111
Qy	661	GGACCTCTTGTTTCCAAATCTGGAATAAACCCATACCTCAAACTTCACCTGTGTAAAAATTT	720
Db	1112		
		GGACCTCTTGTTTCCAAATCTGGAATAAACCCATACCTCAAACTTCACCTGTGTAAAAATTT	1171
Qy	721	AGCAATCTATAGACACAAACAGCTCCCAATGCATCAGGTGGGTAAACACCTCCACACGA	780
Db	1172		
		AGCAATCTATAGACACAAACAGCTCCCAATGCATCAGGTGGGTAAACACCTCCACACGA	1231
Qy	781	ATAGTCTGCCTACCCCTCAGGAATATTTTTGTCTGTGTACTACCTCAGCCTATCATTTGTTG	840
Db	1232		
		ATAGTCTGCCTACCCCTCAGGAATATTTTTGTCTGTGTACTACCTCAGCCTATCATTTGTTG	1291
Qy	841	AATGGCTCTTCAGAAATCTATGTGCTTCTCTCTCAFTTCTTAGTGCCCCCTATGACCATCTAC	900
Db	1292		
		AATGGCTCTTCAGAAATCTATGTGCTTCTCTCTCAFTTCTTAGTGCCCCCTATGACCATCTAC	1351
Qy	901	ACTGAAACAAGATTTATACAATCATGTCGTACTTAAGCCCCACACAAAGAGTACCCATT	960
Db	1352		
		ACTGAAACAAGATTTATACAATCATGTCGTACTTAAGCCCCACACAAAGAGTACCCATT	1411
Qy	961	CTTCTCTTTTGTATCAGACGAGGAGTGCTTAGGCAGACTAGGTACTGGCATTTGGCAGTATC	1020
Db	1412		
		CTTCTCTTTTGTATCAGACGAGGAGTGCTTAGGCAGACTAGGTACTGGCATTTGGCAGTATC	1471
Qy	1021	ACAACTCTTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGTTGATCATGGAAACAG	1080
Db	1472		
		ACAACTCTTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGTTGATCATGGAAACAG	1531
Qy	1081	GTCACTCACTCCCTGGTGCACCTTGCAAGATCAACTTAACCTCCCTAGCAGCAGTAGTCCCT	1140
Db	1532		
		GTCACTCACTCCCTGGTGCACCTTGCAAGATCAACTTAACCTCCCTAGCAGCAGTAGTCCCT	1591
Qy	1141	CAAAATCGAAGAGCTTTAGACTTGTCTAACCGCCAAAAAGGGGAAACCTGTTATTTTAA	1200
Db	1592		
		CAAAATCGAAGAGCTTTAGACTTGTCTAACCGCCAAAAAGGGGAAACCTGTTATTTTAA	1651
Qy	1201	GGAGAAAGACGCTGTTATTATGTTAATCAATCAGAAATGTGCTAGAGAAAGTTAAGAA	1260
Db	1652		
		GGAGAAAGACGCTGTTATTATGTTTAATCAATCCAGAAATGTGCTAGAGAAAGTTAAGAA	1711
Qy	1261	ATTTCGAGATCGAATACAAATGTAGACGAGAGGAGCTTCAAAACACCGAACCTCGGGCCTC	1320
Db	1712		
		ATTTCGAGATCGAATACAAATGTAGACGAGAGGAGCTTCAAAACACCGAACCTCGGGCCTC	1771
Qy	1321	CTCAGCCAAATGGATGCGCTCGGTTCTCCCTCTTCTTAGGACCTCTTAGCAGCTCTAATATTG	1380
Db	1772		
		CTCAGCCAAATGGATGCGCTCGGTTCTCCCTCTTCTTAGGACCTCTTAGCAGCTCTAATATTG	1831
Qy	1381	TTACTCTCTTTTGGACCGCTGTATCTTTAAACCTCTCTGTTAAGTTTGTCTTCCAGAAATT	1440
Db	1832		
		TTACTCTCTTTTGGACCGCTGTATCTTTAAACCTCTCTGTTAAGTTTGTCTTCCAGAAATT	1891

Qy	1441	GAAGCTGTAAGCTACAGATGGTCTTACAAATGGAACCCCA	1481
Db	1892	GAAGCTGTAAGCTACAGATGGTCTTACAAATGGAACCCCA	1932
RESULT 5			
AL390039/c			
LOCUS			
DEFINITION	AL390039	162579 bp	DNA linear PRI 22-DEC-2000
	Human DNA sequence from clone RP13-383K5 on chromosome Xq22.1-24		
	Contains ESTs, STSs and GSSs. Contains a novel gene and a novel		
	gene similar to NUP62 (nucleoporin 62kD), complete sequence.		
ACCESSION	AL390039		
VERSION	AL390039.10	GI:10186780	
KEYWORDS	HTG; nucleoporin; NUP62.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutherii; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 162579)		
AUTHORS	Direct Submission		
TITLE	Submitted (19-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,		
JOURNAL	CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk		
COMMENT	requests: clonesrequest@sanger.ac.uk		
	On Sep 19, 2000 this sequence version replaced gi:9944073.		
	During sequence assembly data is compared from overlapping clones.		
	Where differences are found these are annotated as variations		
	together with a note of the overlapping clone name. Note that the		
	variation annotation may not be found in the sequence submission		
	corresponding to the overlapping clone, as we submit sequences with		
	only a small overlap as described above.		
	The following abbreviations are used to associate primary accession		
	numbers given in the feature table with their source databases:		
	Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information		
	on the WORMPEP database can be found at		
	http://www.sanger.ac.uk/Projects/C.elegans/wormpep/		
	This sequence was generated from part of bacterial clone contigs of human		
	chromosome X, constructed by the Sanger Centre Chromosome X Mapping		
	Group. Further information can be found at		
	http://www.sanger.ac.uk/HGP/ChrX		
	IMPORTANT: This sequence is not the entire insert of clone		
	RP13-383K5 It may be shorter because we sequence overlapping		
	sections only once, except for a 100 base overlap.		
	The true left end of clone RP13-383K5 is at 1 in this sequence. The		
	true left end of clone RP5-1126E12 is at 162480 in this sequence.		
	This sequence has been finished according to sequence map criteria		
	as follows. An attempt is made to resolve all sequencing problems,		
	such as compressions and repeats, but not necessarily within known		
	annotated repeat sequence elements. Where the sequence is		
	ambiguous, there is an annotation using the 'unsure' feature key.		
	RP13-383K5 is from the library RP13-13.2 constructed by the group		
	of Pieter de Jong. For further details see		
	http://www.chori.org/bacpac/home.htm		
	VECTOR: pBACe3.6.		
FEATURES			
source	Location/Qualifiers		
	1..162579		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="X"		
	/map="q22.1-24"		
	/clone="RP13-383K5"		
	/clone_lib="RP13-13.2"		
misc_feature	complement(1..232)		
	/note="match: GSS: Em:AQ036363"		
misc_feature	16..554		
	/note="match: GSS: Em:AQ504848"		
repeat_region	475..644		
	/note="AluSq repeat: matches 134..303 of consensus"		
repeat_region	742..919		
	/note="Mtr1B repeat: matches 207.388 of consensus"		
repeat_region	1112..1142		
	/note="L1MEC repeat: matches 327..356 of consensus"		

```
repeat_region 1143. .1456
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region 1457. .2938
/note="L1MEC repeat: matches 356. .1897 of consensus"
repeat_region 2922. .3739
/note="L1PA2 repeat: matches 5329. .6146 of consensus"
repeat_region 3844. .4087
/note="L1M4 repeat: matches 1. .251 of consensus"
repeat_region 4092. .4231
/note="L1M4 repeat: matches 612. .756 of consensus"
repeat_region 4373. .4559
/note="L1M4c repeat: matches 971. .1152 of consensus"
repeat_region 4586. .4619
/note="17 copies 2 mer aa 100% conserved"
repeat_region 4623. .4964
/note="MLT1A1 repeat: matches 3. .333 of consensus"
repeat_region 4989. .5149
/note="L1MA8 repeat: matches 6120. .6288 of consensus"
repeat_region 5152. .5602
/note="L1PA15 repeat: matches 5325. .5779 of consensus"
repeat_region 5606. .5916
/note="AluSq repeat: matches 1. .313 of consensus"
repeat_region 5917. .6014
/note="49 copies 2 mer ga 75% conserved"
repeat_region 6018. .6409
/note="L1PA15 repeat: matches 5769. .6157 of consensus"
repeat_region 6586. .6750
/note="L1M4 repeat: matches 2857. .3036 of consensus"
repeat_region 6751. .7063
/note="AluJo repeat: matches 1. .307 of consensus"
repeat_region 7064. .8097
/note="L1M4 repeat: matches 3036. .4105 of consensus"
repeat_region 8099. .8419
/note="AluX repeat: matches 1. .306 of consensus"
repeat_region 8420. .8487
/note="L1M4/D repeat: matches 5582. .5648 of consensus"
repeat_region 10039. .10396
/note="L1M4 repeat: matches 1. .364 of consensus"
repeat_region 10519. .10807
/note="AluSq repeat: matches 3. .297 of consensus"
repeat_region 11343. .11416
/note="37 copies 2 mer tt 67% conserved"
repeat_region 11448. .11983
/note="L1MD3 repeat: matches 7162. .7739 of consensus"
repeat_region 12030. .12337
/note="AluX repeat: matches 1. .309 of consensus"
repeat_region 12995. .13077
/note="MIR repeat: matches 103. .186 of consensus"
repeat_region 13110. .13222
/note="FLAM C repeat: matches 2. .124 of consensus"
repeat_region 13228. .13518
/note="L1MEC repeat: matches 2131. .2414 of consensus"
repeat_region 13519. .13853
/note="MER7A repeat: matches 1. .346 of consensus"
repeat_region 13854. .14039
/note="L1MBC repeat: matches 1951. .2131 of consensus"
repeat_region 14040. .14254
/note="MER7A repeat: matches 1. .218 of consensus"
repeat_region 14255. .14807
/note="L1MEC repeat: matches 1686. .1951 of consensus"
repeat_region 14808. .15105
/note="AluB repeat: matches 1. .311 of consensus"
repeat_region 15106. .15492
/note="L1MEC repeat: matches 1309. .1686 of consensus"
repeat_region 15496. .15822
/note="L1MBC repeat: matches 5823. .6167 of consensus"
repeat_region 15824. .15920
/note="L1MBC repeat: matches 1044. .1140 of consensus"
misc_feature 15885. .16086
/note="match: STS: Em:G34444 Em:L14259"
misc_feature 15920. .16082
/note="match: GSS: Em:AQ769824"
misc_feature complement(15927. .16073)

/note="match: GSS: Em:AQ054794 Em:AQ823559"
16067. .16294
/note="match: GSS: Em:AQ457003"
16923. .16998
/note="38 copies 2 mer at 71% conserved"
17107. .17166
/note="L2 repeat: matches 2688. .2750 of consensus"
17175. .17755
/note="L2 repeat: matches 1244. .1898 of consensus"
17839. .18106
/note="MER70A repeat: matches 148. .418 of consensus"
18213. .18318
/note="L2 repeat: matches 1128. .1247 of consensus"
18706. .18811
/note="L2 repeat: matches 2631. .2750 of consensus"
18812. .19057
/note="AluX repeat: matches 1. .297 of consensus"
19058. .19492
/note="L2 repeat: matches 2213. .2647 of consensus"
complement(19352. .19843)
/note="match: GSS: Em:AQ340313"
19556. .19872
/note="L1R17 repeat: matches 1. .326 of consensus"
19873. .21594
/note="HERV17 repeat: matches 6812. .8523 of consensus"
21587. .22280
/note="HERV17 repeat: matches 4301. .4980 of consensus"
22298. .22742
/note="L1MBS repeat: matches 5623. .6135 of consensus"
22743. .23048
/note="AluY repeat: matches 1. .306 of consensus"
23049. .23286
/note="L1MBS repeat: matches 5403. .5623 of consensus"
23285. .23361
/note="L1M4 repeat: matches 5006. .5082 of consensus"
23364. .23429
/note="L1M4 repeat: matches 4019. .4084 of consensus"
23429. .24305
/note="L1 repeat: matches 3097. .4043 of consensus"
24306. .24593
/note="AluJo repeat: matches 1. .293 of consensus"
24594. .25109
/note="L1 repeat: matches 2600. .3097 of consensus"
25110. .25436
/note="L1M3 repeat: matches 5101. .5440 of consensus"
25437. .25742
/note="AluSq repeat: matches 1. .307 of consensus"
25779. .25961
/note="MIR repeat: matches 55. .260 of consensus"
25962. .26059
/note="MIR repeat: matches 1. .89 of consensus"
26135. .26328
/note="MLT1A1 repeat: matches 144. .352 of consensus"
26554. .26697
/note="MLT1A1 repeat: matches 1. .149 of consensus"
26711. .26936
/note="L1PA9 repeat: matches 5939. .6162 of consensus"
26981. .27154
/note="L1MC/D repeat: matches 5620. .5805 of consensus"
27182. .27307
/note="L1P4 repeat: matches 5454. .5584 of consensus"

Query Match 99.1%; Score 1468.2; DB 9; Length 162579;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1473; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ATGCCCTCCCTATCATCTTTCTTTACTGTTCTTACCCCTTCCTCACT 60
Db 21559 ATGCCCTCCCTATCATCTTTCTTTACTGTTCTTACCCCTTCCTCACT 21500
Qy 61 GCACCCCTCATGCTGCTGTATCAACCAAGTAGTCCCTTACCAAGAGTTCTATGAAGA 120
Db 21499 GCACCCCTCATGCTGCTGTATCAACCAAGTAGTCCCTTACCAAGAGTTCTATGAAGA 21440
```

QY 121 ACAGGCTTCTGGAATATTTGATGCCCCATCATATAGAGTTTATCTAAGGGAATCTCC 180
DB 21439 ACAGGCTTCTGGAATATTTGATGCCCCATCATATAGAGTTTATCTAAGGGAATCTCC 21380
QY 181 ACCTTCACTGCCACACCCATATGCCCCGAACTGCTATAAATCTGCACTCTTTGCATG 240
DB 21379 ACCTTCACTGCCACACCCATATGCCCCGAACTGCTATAAATCTGCACTCTTTGCATG 21320
QY 241 CATGCAATACTCATTTATGGACAGGGAATAATGATTAATCCTAGTTGTCTCGAGGACTT 300
DB 21319 CATGCAATACTCATTTATGGACAGGGAATAATGATTAATCCTAGTTGTCTCGAGGACTT 21260
QY 301 GGAGCCACTGTCTGTTGGACTTACTCACCATACAGATATGCTGATGGGGTGAATTT 360
DB 21259 GGAGCCACTGTCTGTTGGACTTACTCACCATACAGATATGCTGATGGGGTGAATTT 21200
QY 361 CAAAGGTTCAGGCAAGAGAAAAAACAAGTAAGGAAGCAATCTCCCACTGACCCGGGACAT 420
DB 21199 CAAAGGTTCAGGCAAGAGAAAAAACAAGTAAGGAAGCAATCTCCCACTGACCCGGGACAT 21140
QY 421 AGCACCCCTAGCCCTACAAAGGACTAGTTCTCTCAAACTACATGAAACCCCTCCGTACC 480
DB 21139 AGCACCCCTAGCCCTACAAAGGACTAGTTCTCTCAAACTACATGAAACCCCTCCGTACC 21080
QY 481 CATACTCGCTGTGAGCCTATTATTAACACCCCTCACTCGGCTCCATGAGGTCTCAGCC 540
DB 21079 CATACTCGCTGTGAGCCTATTATTAACACCCCTCACTCGGCTCCATGAGGTCTCAGCC 21020
QY 541 CAAAACCCCTACTAACTGTTGGATGTGCTCCCTCGACCTTCAAGGCATACATTTCAATC 600
DB 21019 CAAAACCCCTACTAACTGTTGGATGTGCTCCCTCGACCTTCAAGGCATACATTTCAATC 20960
QY 601 CTGTTTCTGAAACAATGGAACAATCTCAGCACAGAAATAAACACCTTCCGTTTGTAGTA 660
DB 20959 CTGTTTCTGAAACAATGGAACAATCTCAGCACAGAAATAAACACCTTCCGTTTGTAGTA 20900
QY 661 GGACCTTCTGTTTCCAACTCGGAATAAACCATACTCAAACTCACTGTTGTAATAATTT 720
DB 20899 GGACCTTCTGTTTCCAACTCGGAATAAACCATACTCAAACTCACTGTTGTAATAATTT 20840
QY 721 AGCAATACTATAGACACACACCTCCCAATGATCAGGTGGGTAAACCTCCACACGA 780
DB 20839 AGCAATACTATAGACACACACCTCCCAATGATCAGGTGGGTAAACCTCCACACGA 20780
QY 781 ATAGTCTGCTTACCTCAGGAATATTTTGTCTGTGTACCTCAGCCTATCATTTGTTG 840
DB 20779 ATAGTCTGCTTACCTCAGGAATATTTTGTCTGTGTACCTCAGCCTATCATTTGTTG 20720
QY 841 AATGGCTTTCAGAACTATGTCCTTCTCATTTCTTAGTGCCTTATGACCATCTAC 900
DB 20719 AATGGCTTTCAGAACTATGTCCTTCTCATTTCTTAGTGCCTTATGACCATCTAC 20660
QY 901 ACTGAACAGATTATATACATCATGTCGTACCTAAGCCCCCAACAAAGAGTACCATT 960
DB 20659 ACTGAACAGATTATATACATCATGTCGTACCTAAGCCCCCAACAAAGAGTACCATT 20600
QY 961 CTTCTCTTTTGTATCAGAGCAGAGTGTAGGAGAGTACTAGGCTAGTCTGTCATGTCAGTATC 1020
DB 20599 CTTCTCTTTTGTATCAGAGCAGAGTGTAGGAGAGTACTAGGCTAGTCTGTCATGTCAGTATC 20540
QY 1021 ACAACCTTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGGTGAATGGAACAG 1080
DB 20539 ACAACCTTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGGTGAATGGAACAG 20480
QY 1081 GTCACTGACTCCCTGTGCTTCAAGATCACTTAATCTCCCTAGCAGGAGTCCCTT 1140
DB 20479 GTCACTGACTCCCTGTGCTTCAAGATCACTTAATCTCCCTAGCAGGAGTCCCTT 20420
QY 1141 CAAAATCGAAGAGCTTTAGACTTTGCTAAACCGCAAAAGAGGGGGAACCTGTTTATTTTAA 1200
DB 20419 CAAAATCGAAGAGCTTTAGACTTTGCTAAACCGCAAAAGAGGGGGAACCTGTTTATTTTAA 20360

QY 1201 GGAGRAAGCGCTGTTATTATGTTAATCAATCCAGAAATTTGCTCACTGAGRAAGTTAAGAA 1260
DB 20359 GGAGRAAGATGCTGTTATTATGTTAATCAATCCAGAAATTTGCTCACTGAGRAAGTTAAGAA 20300
QY 1261 ATTCGAGATCGGAATACAAATGTAGAGCAGAGGAGCTTCAAAACACCGAACCTCGGGCCTC 1320
DB 20299 ATTCGAGATCGGAATACAAATGTAGAGCAGAGGAGCTTCAAAACACCGAACCTCGGGCCTC 20240
QY 1321 CTCAGCAATGAGATGCGCTGCTTCCCTTCTTAGGACCTCTAGCAGCTCTAAATATTG 1380
DB 20239 CTCAGCAATGAGATGCGCTGCTTCCCTTCTTAGGACCTCTAGCAGCTCTAAATATTG 20180
QY 1381 TTACTCTCTTTGGACCTGTATCTTAACCTCTTGTAAAGTTTCTCTCCAGAAAT 1440
DB 20179 TTACTCTCTTTGGACCTGTATCTTTAACTCTTGTAAAGTTTCTCTCCAGAAAT 20120
QY 1441 GAAGCTGTAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
DB 20119 GAAGCTGTAAGCTACAAATGGTCTTCAATGGAGCCCA 20079

RESULT 6
LOCUS CQ719381 2055 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 5315 from Patent WO02068579.
ACCESSION CQ719381
VERSION CQ719381.1 GI:42280238
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof
JOURNAL Patent: WO 02068579-A 5315 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
Location/Qualifiers
1..2055
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 99.0%; Score 1466.6; DB 6; Length 2055;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1472; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ATGGCCCTCCCTTATCATACTTTTCTTTTACTGTTCTCTTACCCCTTTTCGCTCTCACT 60
DB 36 ATGGCCCTCCCTTATCATATTTTCTTTTACTGTTCTCTTACCCCTTTTCGCTCTCACT 95
QY 61 GCACCCCTCCATGCTGCTGTACAACAGTAGTCCCTTACCAAGATTCTTATGAGA 120
DB 96 GCACCCCTCCATGCTGCTGTACAACAGTAGTCCCTTACCAAGATTCTTATGAGA 155
QY 121 AGCGGCTTCTCGGAATATTTGATGCCCATATATAGAGTTTATCTAAGGGAATCTCC 180
DB 156 ACAGGCTTCTCGGAATATTTGATGCCCATATATAGAGTTTATCTAAGGGAATCTCC 215
QY 181 ACCTTCACTGCCACACCCATATGCCCCGAACTGCTATAAATCTCTTTCATG 240
DB 216 ACCTTCACTGCCACACCCATATGCCCCGAACTGCTATAAATCTCTTTCATG 275
QY 241 CATGCAATACTCATTTATGGACAGGGAATAATGATTAATCCTAGTTGTCTCGAGACTT 300
DB 276 CATGCAATACTCATTTATGGACAGGGAATAATGATTAATCCTAGTTGTCTCGAGACTT 335
QY 301 GGAGCCACTGTCTGTTGGACTTACTCACCATACAGATATGCTGATGGGGTGAATTT 360
DB 336 GGAGCCACTGTCTGTTGGACTTACTCACCATACAGATATGCTGATGGGGTGAATTT 395

QY 361 CAAGGTCAGGCAAGAGAAACAAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGACAT 420
Db 396 CAAGGTCAGGCAAGAGAAACAAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGACAT 455
QY 421 AGCACCCCTAGCCCTACAAAGACATAGTTCTCTCAAACTACATGAAACCCCTCCGTACC 480
Db 456 AGCACCCCTAGCCCTACAAAGACATAGTTCTCTCAAACTACATGAAACCCCTCCGTACC 515
QY 481 CATACTCGCTGTGAGCCCTATTATTAATACCACTCAGTCGGCTCCATGAGGCTCAGCC 540
Db 516 CATACTCGCTGTGAGCCCTATTATTAATACCACTCAGTCGGCTCCATGAGGCTCAGCC 575
QY 541 CAAAACCTACTAACTGTGCGATGTGCTCCCTCCCTGCATCTCAGGCGCATACATTTCAATC 600
Db 576 CAAAACCTACTAACTGTGCGATGTGCTCCCTCCCTGCATCTCAGGCGCATACATTTCAATC 635
QY 601 CTTGTTCTGAACAATGGAACAACTTCAGCACAGAAATAAACAACCACTCCCGTTTAGTA 660
Db 636 CTTGTTCTGAACAATGGAACAACTTCAGCACAGAAATAAACAACCACTCCCGTTTAGTA 695
QY 661 GGACCTCTGTTTCCCAATCTGGAATAACCATACCTCRAACCTCAGCTGTGTAATAATT 720
Db 696 GGACCTCTGTTTCCCAATCTGGAATAACCATACCTCRAACCTCAGCTGTGTAATAATT 755
QY 721 AGCAATACTATAGACAAACAGCTCCCAATGCATCAGTGGGTAAACACCTCCACACGA 780
Db 756 AGCAATACTATAGACAAACAGCTCCCAATGCATCAGTGGGTAAACACCTCCACACGA 815
QY 781 ATAGTCTGCTACCCCTCAGGAATATTTTTGTCTGTGGTACCTCAGGCTATCATTTGTTG 840
Db 816 ATAGTCTGCTACCCCTCAGGAATATTTTTGTCTGTGGTACCTCAGGCTATCATTTGTTG 875
QY 841 AATGGCTCTTCAGAACTATGTGCTCCTCTCATCTTAGTGCCCTATGACCATCTAC 900
Db 876 AATGGCTCTTCAGAACTATGTGCTCCTCTCATCTTAGTGCCCTATGACCATCTAC 935
QY 901 ACTGAACAAGATTATACAACTCATGTCTGTAAGCCCAACCAACAAAGAGTACCCATT 960
Db 936 ACTGAACAAGATTATACAACTCATGTCTGTAAGCCCAACCAACAAAGAGTACCCATT 995
QY 961 CTTCCCTTTGTTATCAGACGAGGAGTGTCTAGGAGACTAGGTACTGCGATGGCAGTATC 1020
Db 996 CTTCCCTTTGTTATCAGACGAGGAGTGTCTAGGAGACTAGGTACTGCGATGGCAGTATC 1055
QY 1021 ACACCTCTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
Db 1056 ACACCTCTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1115
QY 1081 GTCACCTGACCTCCCTGGTCACTTGCAGATCACTTAACCTCCCTAGCAGCAGTAGTCCTT 1140
Db 1116 GTCACCTGACCTCCCTGGTCACTTGCAGATCACTTAACCTCCCTAGCAGCAGTAGTCCTT 1175
QY 1141 CAAAATCGAAGAGCTTTAGACTTGTCTAACCCGCAAGAGGGGAACTGTTTATTTTAA 1200
Db 1176 CAAAATCGAAGAGCTTTAGACTTGTCTAACCCGCAAGAGGGGAACTGTTTATTTTAA 1235
QY 1201 GGAGAAGAACCTGTTATTTATGTTAACTCAATCCAGAAATTTGTCATGAGAAAGTTAAAGAA 1260
Db 1236 GGAGAAGAACCTGTTATTTATGTTAACTCAATCCAGAAATTTGTCATGAGAAAGTTAAAGAA 1295
QY 1261 ATTGAGATCGAATACAAATGTAGACGAGAGGCTTCAAAACACCGAAGCTGGGGCTC 1320
Db 1296 ATTGAGATCGAATACAAATGTAGACGAGAGGCTTCAAAACACCGAAGCTGGGGCTC 1355
QY 1321 CTCAGCAATGGATGCCCTGGGTTCTCCCTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
Db 1356 CTCAGCAATGGATGCCCTGGGTTCTCCCTCTTAGGACCTCTAGCAGCTCTAATATTG 1415
QY 1381 TTACTCTCTTTGGACCCCTGTATCTTTAACTCTCTTTAAAGTTTGTCTCTCCAGAAAT 1440
Db 1416 TTACTCTCTTTGGACCCCTGTATCTTTAACTCTCTTTAAAGTTTGTCTCTCCAGAAAT 1475

QY 1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAAACCCCA 1481
Db 1476 GAAGCTGTAAAGCTACAAATGGTCTTCAAAATGGAGCCCA 1516
RESULT 7
BD252040 1629 bp DNA linear PAT 17-JUL-2003
LOCUS Method for detecting superantigen activity in a biological sample.
DEFINITION
ACCESSION BD252040
VERSION BD252040.1 GI:33061810
KEYWORDS JP 2002539804-A/1
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1629)
AUTHORS Perron H. and Lafont, M.
TITLE Method for detecting superantigen activity in a biological sample
JOURNAL Patent: JP 2002539804-A 1 26-NOV-2002;
BIO MERIEUX
COMMENT OS Homo sapiens (human)
PN JP 2002539804-A/1
PD 26-NOV-2002
PF 20-MAR-2000 JP 2000607009 99/03622, 28-OCT-1999 FR 99/13755 PI
PR 19-MAR-1999 FR 99/03622, 28-OCT-1999 FR 99/13755 PI
PC C12N15/09, A61K31/711, A61K38/00, A61K39/395, A61K45/00, A61K48/00,
A61P25/00,
PC C07K14/47, C07K16/18, C12N7/00, C12Q1/02, C12Q1/68, G01N27/447, PC
G01N33/15,
G01N33/50, G01N33/569, G01N33/577//C12P21/08, (C12N7/00, C12R1:93), PC
(C12Q1/02, C12R1:93), (C12Q1/68, C12R1:93), C12N15/00, G01N27/26, PC
A61K37/02
CC Method for detecting superantigen activity in a biological CC
FEATURES
source
1. .1629
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 98.7%; Score 1461.8; DB 6; Length 1629;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1469; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 1 ATGGCCCTCCCTTATCATACTTTTCTTTTACTTGTCTCTTACCCCTTTCGCTCTCACT 60
Db 1 ATGGCCCTCCCTTATCATACTTTTCTTTTACTTGTCTCTTACCCCTTTCGCTCTCACT 60
QY 61 GCACCCCTCCAGCTGTGTACCAACAGTAGTCCCTTACCAGAGTTTCTATGAAGA 120
Db 61 GCACCCCTCCAGCTGTGTACCAACAGTAGTCCCTTACCAGAGTTTCTATGAAGA 120
QY 121 ACGGCGCTTCTCGAAATATTGATGCCCATCATATAGGAGTTTATTAAGGAAACTCC 180
Db 121 ACGGCGCTTCTCGAAATATTGATGCCCATCATATAGGAGTTTATTAAGGAAACTCC 180
QY 181 ACCTTCACTGCCACACCCATATGCCCGCAATGCTATTAACCTGCTCCACTCTTTCATG 240
Db 181 ACCTTCACTGCCACACCCATATGCCCGCAATGCTATTAACCTGCTCCACTCTTTCATG 240
QY 241 CATGCAATACTCATTTATGACAGGAAATGATTAATCCTAGTTGTCTCGGAGGACTT 300
Db 241 CATGCAATACTCATTTATGACAGGAAATGATTAATCCTAGTTGTCTCGGAGGACTT 300
QY 301 GGAGCCACTGTCTGTGGACTTACTTCCACCATACAGTATGCTGATGGGGTGGAAAT 360
Db 301 GGAGCCACTGTCTGTGGACTTACTTCCACCATACAGTATGCTGATGGGGTGGAAAT 360

301 GGAGCCACTGCTCTGTTGGACTTACTTACCCATACACAGTATGCTCTGATGGGGTGGAAATT 360
361 CAAGGTCAGCAAGAGAAAACAAGTAAGGAAGCAATCTCCCACTGACCCCGGGACAT 420
361 CAAGGTCAGCAAGAGAAAACAAGTAAGGAAGCAATCTCCCACTGACCCCGGGACAT 420
421 AGCACCCCTAGCCCTACAAAGGACTAGTTCTCTCAAACTACATGAAACCCCTCCGTACC 480
421 AGCACCCCTAGCCCTACAAAGGACTAGTTCTCTCAAACTACATGAAACCCCTCCGTACC 480
481 CATACTCGCTGCTGAGCCCTATTTAATACCAACCCCTCACTCGGCTCCATGAGGTCTCAGCC 540
481 CATACTCGCTGCTGAGCCCTATTTAATACCAACCCCTCACTCGGCTCCATGAGGTCTCAGCC 540
541 CAAAACCCCTACTAACTCTGAGTGGTCTCCCTCCCTGACCTTTCAGGCATACATTTCAATC 600
541 CAAAACCCCTACTAACTCTGAGTGGTCTCCCTCCCTGACCTTTCAGGCATACATTTCAATC 600
601 CTGTTCTCTGAACAATGGAACTCTCAGCAGAAATAAACACCCACTTCCCGTTTGTAGTA 660
601 CTGTTCTCTGAACAATGGAACTCTCAGCAGAAATAAACACCCACTTCCCGTTTGTAGTA 660
661 GGACCTCTTGTCTTCCCAATCTGGAATAAACCCTACCTCAAACTCACCCTGTGTAAAATT 720
661 GGACCTCTTGTCTTCCCAATCTGGAATAAACCCTACCTCAAACTCACCCTGTGTAAAATT 720
721 AGCAATCTATAGACACCAACAGCTCCCAATGATCAGTGGGTAAACACCTCCACACGA 780
721 AGCAATCTATAGACACCAACAGCTCCCAATGATCAGTGGGTAAACACCTCCACACGA 780
781 ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGTGATCTCAGCCTATCATTTGTTG 840
781 ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGTGATCTCAGCCTATCATTTGTTG 840
841 AATGGCTCTTCAGAACTATGCTCTCTCATCTTACTGCTCCCTATGACCATCTAC 900
841 AATGGCTCTTCAGAACTATGCTCTCTCATCTTACTGCTCCCTATGACCATCTAC 900
901 ACTGAACAAGATTATACAACTATGTCGTAACCTAAGCCCAACAAAGAGTACCCATT 960
901 ACTGAACAAGATTATACAACTATGTCGTAACCTAAGCCCAACAAAGAGTACCCATT 960
961 CTTCTCTTTGTTATCAGACAGAGTGTAGGACAGACTAGGTACTGGCATTGGCAGTATC 1020
961 CTTCTCTTTGTTATCAGACAGAGTGTAGGACAGACTAGGTACTGGCATTGGCAGTATC 1020
1021 ACAACCTCTACTCAGTCTTACTACAACTATCTCAAGAAATAAATGCTGACATGGAAACAG 1080
1021 ACAACCTCTACTCAGTCTTACTACAACTATCTCAAGAAATAAATGCTGACATGGAAACAG 1080
1081 GTCACCTGACTCCCTGGTCTACCTTTGCAAGATCAACTTAACCTCCCTAGCAGCAGTGCCTT 1140
1081 GTCACCTGACTCCCTGGTCTACCTTTGCAAGATCAACTTAACCTCCCTAGCAGCAGTGCCTT 1140
1141 CAAAATCGAAGACTTTAGACTTGTCTAAACCGCAAAAGAGGGGGAACCTGTTTATTTTA 1200
1141 CAAAATCGAAGACTTTAGACTTGTCTAAACCGCAAAAGAGGGGGAACCTGTTTATTTTA 1200
1201 GGAAGAAAGCGCTGTTATTTATGTTAATCAATCCAGAAATGCTCACTGAGAAAGTTAAGAA 1260
1201 GGAAGAAAGCGCTGTTATTTATGTTAATCAATCCAGAAATGCTCACTGAGAAAGTTAAGAA 1260
1261 ATTTCGAGATCGAATACAAATGAGCAGAGAGCTTCAAAACACCGCAAGCTGGGCGCTC 1320
1261 ATTTCGAGATCGAATACAAATGAGCAGAGAGCTTCAAAACACCGCAAGCTGGGCGCTC 1320
1321 CTGAGCAATGGAATGCGCTGGTCTTCCCTCTTCTAGGACCTCTAGAGCTCTAAATTTG 1380
1321 CTGAGCAATGGAATGCGCTGGTCTTCCCTCTTCTAGGACCTCTAGAGCTCTAAATTTT 1380
1381 TTAATCTCTTTGGACCGCTGATCTTTAACTCTCTTTGTTAAGTTTGTCTCTCCAGAAAT 1440
1381 TTAATCTCTTTGGACCGCTGATCTTTAACTCTCTTTGTTAAGTTTGTCTCTCCAGAAAT 1440

QY 1441 GAAGCTGTAAGCTACAGATGGTCTTTACAAATGGAACCCCA 1481
Db 1441 GAAGCTGTAAGCTACAAATAGTTCTTCAATGGAACCCCA 1481

RESULT 8
AX036992

LOCUS AX036992 1629 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 1 from Patent FR2791140.
ACCESSION AX036992
VERSION AX036992.1 GI:11226420
KEYWORDS
SOURCE Homo sapiens (human)

ORGANISM
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 Lafont, M. and Perron, H.
AUTHORS
JOURNAL Patent: FR 2791140-A 1 22-SEP-2000;
BIO MERIEUX (FR)

FEATURES

Location/Qualifiers
source 1..1629
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 98.7%; Score 1461.8; DB 6; Length 1629;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1469; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 ATGGCCCTCCCTTATCATACTTTTCTTTTACTGTTTCTTTACCCCTTTTCGCTCTCACT 60

Db 1 ATGGCCCTCCCTTATCATACTTTTCTTTTACTGTTTCTTTACCCCTTTTCGCTCTCACT 60

QY 61 GCACCCCTCCATGCTGCTGTACCAACAGTAGTCCCTTACCAAGATTCTTATGAGA 120

Db 61 GCACCCCTCCATGCTGCTGTACCAACAGTAGTCCCTTACCAAGATTCTTATGAGA 120

QY 121 ACGCGCTCTCTCGAAATATTTGATGCCCATCATATAGGAGTTTATCTAAGGGAATCC 180

Db 121 ACGCGCTCTCTCGAAATATTTGATGCCCATCATATAGGAGTTTATCTAAGGGAATCC 180

QY 181 ACCTTCACTGCCACACCCCATATGCCCCGCAACTGCTATAAATCTCTTGGCATG 240

Db 181 ACCTTCACTGCCACACCCCATATGCCCCGCAACTGCTATAAATCTCTTGGCATG 240

QY 241 CATGCAATCTCATTTATTTGGACAGGGAATGATTAATCTAGTTGTCCTGGAGGACTT 300

Db 241 CATGCAATCTCATTTATTTGGACAGGGAATGATTAATCTAGTTGTCCTGGAGGACTT 300

QY 301 GGAGCCACTGCTGTTGGACTTACTTACCCTACAGTATGCTCATGGGGTGGAAAT 360

Db 301 GGAGCCACTGCTGTTGGACTTACTTACCCTACAGTATGCTCATGGGGTGGAAAT 360

QY 361 CAAGGTGAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGACAT 420

Db 361 CAAGGTGAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGACAT 420

QY 421 AGCACCCCTAGCCCTTACAAAGGACTAGTTCTCTCAAACTACATGAACCCCTCCGTACC 480

Db 421 AGCACCCCTAGCCCTTACAAAGGACTAGTTCTCTCAAACTACATGAACCCCTCCGTACC 480

QY 481 CATACTCGCTGCTGAGCCTTATTAATACCCCTCACTCGGCTCCATGAGGTCTCAGCC 540

Db 481 CATACTCGCTGCTGAGCCTTATTAATACCCCTCACTCGGCTCCATGAGGTCTCAGCC 540

QY 541 CAAAACCCCTACTAACTGTTGGATGTCCTCCCTCGACTTTCAGGCCATACATTTCAATC 600

Db 541 CAAAACCCCTACTAACTGTTGGATGTCCTCCCTCGACTTTCAGGCCATACATTTCAATC 600

QY 601 CCTGTTCTTGAAATGGAACAACTTTCAGCACAGAAATAAACACCACTCTCCGTTTGTAGTA 660

Db 601 CCTGTTCTCTGAAACAATGGAACACTTTCAGCACAGAAATAAACACCACTTCCGGTTTATGTA 660
Qy 661 GGACCTCTTGTTCCTCAATCTGGAATAAACCCATACCTCAACCTACCTGCTGTGTAATTT 720
Db 661 GGACCTCTTGTTCCTCAATCTGGAATAAACCCATACCTCAACCTACCTGCTGTGTAATTT 720
Qy 721 AGCAATACTATAGACACACACCTCCCAATGATCAGGTGGGTGAACACCTCCACACCA 780
Db 721 AGCAATACTATAGACACACACCTCCCAATGATCAGGTGGGTGAACACCTCCACACCA 780
Qy 781 ATAGTCTGCTACCTCCAGGAATATTTTGTCTGCTGATCCTCAACCTCAATCTGTTTG 840
Db 781 ATAGTCTGCTACCTCCAGGAATATTTTGTCTGCTGATCCTCAACCTCAATCTGTTTG 840
Qy 841 AATGGCTCTCAGAACTATGCTGCTCTCTCAATCTTAGTGCCCTATGACCACTAC 900
Db 841 AATGGCTCTCAGAACTATGCTGCTCTCTCAATCTTAGTGCCCTATGACCACTAC 900
Qy 901 ACTGAACAAGATTATACAACTATGCTGATCCTTAAGCCCAACAAAGAGTACCCATT 960
Db 901 ACTGAACAAGATTATACAACTATGCTGATCCTTAAGCCCAACAAAGAGTACCCATT 960
Qy 961 CTTCTCTTTGTTATCAGACAGAGTGTCTAGGAGACTAGGTACTGSCATTTGCGAGTATC 1020
Db 961 CTTCTCTTTGTTATCAGACAGAGTGTCTAGGAGACTAGGTACTGSCATTTGCGAGTATC 1020
Qy 1021 ACAACCTCTACTAGTCTACTACAACTATCTCAGAAATAAATGCTGACATGGAACAG 1080
Db 1021 ACAACCTCTACTAGTCTACTACAACTATCTCAGAAATAAATGCTGACATGGAACAG 1080
Qy 1081 GTCACTGACCTCCCTGGTCACTTGCAGAGTCAACTTAACCTCCCTAGCAGAGTAGTCCCT 1140
Db 1081 GTCACTGACCTCCCTGGTCACTTGCAGAGTCAACTTAACCTCCCTAGCAGAGTAGTCCCT 1140
Qy 1141 CAAAATCGAAGAGCTTTAGACTTGTCTAACCCGCAAAAGAGGGGAACTCTTTATTTTA 1200
Db 1141 CAAAATCGAAGAGCTTTAGACTTGTCTAACCCGCAAAAGAGGGGAACTCTTTATTTTA 1200
Qy 1201 GGAGAAGAACGCTGTTATTTATGTTAATCAATCCAGATTCCTCACTGAGAAAGTTAAAGAA 1260
Db 1201 GGAGAAGAACGCTGTTATTTATGTTAATCAATCCAGATTCCTCACTGAGAAAGTTAAAGAA 1260
Qy 1261 ATTGAGATCGAATACAACTATGAGCAGAGAGCTTCAAAACACCGAAGCTGGGGCTC 1320
Db 1261 ATTGAGATCGAATACAACTATGAGCAGAGAGCTTCAAAACACCGAAGCTGGGGCTC 1320
Qy 1321 CTCAGCAATGGATGGCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
Db 1321 CTCAGCAATGGATGGCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
Qy 1381 TTACTCTCTTTGACCTGATCTTTAACTCCCTCTGTTAAGTTTGTCTCTCCAGAAAT 1440
Db 1381 TTACTCTCTTTGACCTGATCTTTAACTCCCTCTGTTAAGTTTGTCTCTCCAGAAAT 1440
Qy 1441 GAAGCTGTAAGCTACAAATAGTTCTTCAATGGAACCCCA 1481
Db 1441 GAAGCTGTAAGCTACAAATAGTTCTTCAATGGAACCCCA 1481

RESULT 9
AF331500
LOCUS AF331500
DEFINITION Multiple sclerosis associated retrovirus element clone pv14
ACCESSION AF331500
VERSION AF331500
KEYWORDS AF331500.1 GI:13310190
SOURCE Multiple sclerosis associated retrovirus element
ORGANISM Multiple sclerosis associated retrovirus element
REFERENCE 1 (bases 1 to 1629)
AUTHORS Perron,H., Jouvin-Marche,E., Michel,M., Ounanian-Paraz,A.,

Camelo,S., Dumon,A., Jolivet-Reyraud,C., Marcel,F., Souillet,Y.,
Borel,E., Gebuhrer,L., Santoro,L., Marcel,S., Seigneurin,J.M.,
Marche,P.N. and Lafon,M.
Multiple sclerosis retrovirus particles and recombinant envelope
trigger an abnormal immune response in vitro, by inducing
polyclonal Vbeta16 T-lymphocyte activation
Virology 287 (2), 321-332 (2001)
21425392
11531410
2 (bases 1 to 1629)
Perron H., Ounanian-Paraz,A., Marcel,F. and Souillet,Y.
Direct Submission
Submitted (21-DEC-2000) R&D, bioMerieux, Chemin de L'Orme, Marcy
L'Etoile 69280, France
Location/Qualifiers
FEATURES
1..1629
/organism="Multiple sclerosis associated retrovirus
element"
/mol_type="mRNA"
/db_xref="taxon:89382"
/clone="pv14"
1..1629
/gene="env"
1..1629
/gene="env"
/function="displaying superantigen-like properties causing
polyclonal activation of human Vbeta 16 T-cells (CD3+)
independent from the T-cell receptor antigen specificity"
/note="surface and transmembrane regions"
/codon_start=1
/product="recombinant envelope protein"
/protein_id="AAK18189.1"
/db_xref="GI:13310191"
/translation="MALPYHTFLFTVLLPFPALTAPPPCCCTSSPYQEFUWRTLP
GNIDAPSVRLSKGNSFTTAHTMPKNCYNATLCMHANTHYWGMNINPSCPGGLG
TWCTVYTHTSMDGGIOGAAREKOVKEAISOLTRGHTSPGPKGLVLSKLEHILRT
HRLVSLFNTTLRLHEVSAQNPNCMLPLHFRPVIISIPVEQNNSTFENTTSV
LVGPVNSUEITHSNLTCVKFSNTLIDTSSQCIKRWTPPTRIVCLPSGIFVCGTSA
YHCLNGSSSMCMFLPVPMTTYTEQDLYNHVPKPNKRVLFVIRAGVILGRIG
TGISTTLTQFYKLSQINGMEQVTLSDIOLQNSLAALVIONRAUULLLTK
RGCTCLFLGEERYVYVNSRIKVEKRIQCRAEELONTERGLLSQMMWPTLP
FLGPLAAIIFLLFGPCIENFLVKKFVSSRIEAVKLQIVLQMEPQMSMTKIYRPLDR
PARLSDVNDIEVTPPEEISTAQPLHNSVSGSS"
ORIGIN
Query Match 98.7%; Score 1461.8; DB 14; Length 1629;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1469; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
Qy 1 ATGGCCCTCCCTTATCATATCTTTTCTTTACTGTTCTTTACCCCTTTTCGCTCTCACT 60
Db 1 ATGGCCCTCCCTTATCATATCTTTTCTTTACTGTTCTTTACCCCTTTTCGCTCTCACT 60
Qy 61 GCACCCCTCCATGCTGTCAACACAGTAGTCCCTTTACCAAGAGTTTCTATGAAGA 120
Db 61 GCACCCCTCCATGCTGTCAACACAGTAGTCCCTTTACCAAGAGTTTCTATGAAGA 120
Qy 121 ACGCGGCTTCTCGAAATATTGATGCCCCCATATAGGAGTTTATTAAGGAAATCC 180
Db 121 ACGCGGCTTCTCGAAATATTGATGCCCCCATATAGGAGTTTATTAAGGAAATCC 180
Qy 181 ACCTTCACTGCCACACCCATATGCCCGCACTGCTATTAACCTCTTTCATG 240
Db 181 ACCTTCACTGCCACACCCATATGCCCGCACTGCTATTAACCTCTTTCATG 240
Qy 241 CATGCAAACTACTCATTTATGACAGGGAATAATGATTAATCTAGTTCTCTGGAGGACTT 300
Db 241 CATGCAAACTACTCATTTATGACAGGGAATAATGATTAATCTAGTTCTCTGGAGGACTT 300
Qy 301 GGAGCCACTGCTGTGTTGGACTTACTTCAACCCATACAGTATGTTGATGGGGTGGAAAT 360
Db 301 GGAGCCACTGCTGTGTTGGACTTACTTCAACCCATACAGTATGTTGATGGGGTGGAAAT 360

361 CAAGGTCAGGCAAGAAAACAAAGTAAGAAAGCAATCTCCAACTGACCCCGGACAT 420
Db 361 CAAGGTCAGGCAAGAAAACAAAGTAAGAAAGCAATCTCCAACTGACCCCGGACAT 420
QY 421 AGCACCCCTAGCCCTACAAAGGAGTAGTTCTCTCAAAAATACATGAAACCCCTCCGTACC 480
Db 421 AGCACCCCTAGCCCTACAAAGGAGTAGTTCTCTCAAAAATACATGAAACCCCTCCGTACC 480
QY 481 CATATCGCCTGGTGAGCCCTATTATATACACCCCTCACTCGGCTCCATGAGGTCAGCC 540
Db 481 CATATCGCCTGGTGAGCCCTATTATATACACCCCTCACTCGGCTCCATGAGGTCAGCC 540
QY 541 CAAAACCCCTACTACTGTCGATGTCCTCCCTGCACCTTCAGGCCATACATTCATC 600
Db 541 CAAAACCCCTACTACTGTCGATGTCCTCCCTGCACCTTCAGGCCATACATTCATC 600
QY 601 CTGTGTCCTGAAACAATGGAACAACTTCAGCACAGAAATAAACCACTTCGCTTTAGTA 660
Db 601 CTGTGTCCTGAAACAATGGAACAACTTCAGCACAGAAATAAACCACTTCGCTTTAGTA 660
QY 661 GGACCTCTGTTTCCAACTTGGAAATAAACCCATACCTCAAACTCACCTGTGTAATTT 720
Db 661 GGACCTCTGTTTCCAACTTGGAAATAAACCCATACCTCAAACTCACCTGTGTAATTT 720
QY 721 AGCAATACTATAGACACCAAGCTCCCAATGCATCAGGTGGGTAAACACCTCCACACGA 780
Db 721 AGCAATACTATAGACACCAAGCTCCCAATGCATCAGGTGGGTAAACACCTCCACACGA 780
QY 781 ATAGTCTGCTTACCTCAGCAATATTTTGTCTGTGTACTCAGGCTATCATTTGTTG 840
Db 781 ATAGTCTGCTTACCTCAGCAATATTTTGTCTGTGTACTCAGGCTATCATTTGTTG 840
QY 841 AATGGCTCTTTCAGAACTATGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
Db 841 AATGGCTCTTTCAGAACTATGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
QY 901 ACTGAACAAGATTATATCAATCATGCTGCTACCTAAGCCCAACAAAGAGTACCCATT 960
Db 901 ACTGAACAAGATTATATCAATCATGCTGCTACCTAAGCCCAACAAAGAGTACCCATT 960
QY 961 CTTTCTTTTCTTATCAGACAGAGTGCTAGGAGACTAGGTACTGTGCTGCTGCTGCTGCT 1020
Db 961 CTTTCTTTTCTTATCAGACAGAGTGCTAGGAGACTAGGTACTGTGCTGCTGCTGCTGCT 1020
QY 1021 ACAACCTCTACTCAGTTTCTACTCAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
Db 1021 ACAACCTCTACTCAGTTTCTACTCAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
QY 1081 GTCACTGACTCCCTGGTCACTTGCAGATCAACTTAACCTCCCTAGCAGCAGTAGTCCTT 1140
Db 1081 GTCACTGACTCCCTGGTCACTTGCAGATCAACTTAACCTCCCTAGCAGCAGTAGTCCTT 1140
QY 1141 CAAAATCGAAGAGCTTTAGACTTGTAAACCGCAAGAGGGGGAACCTGTTTATTTTA 1200
Db 1141 CAAAATCGAAGAGCTTTAGACTTGTAAACCGCAAGAGGGGGAACCTGTTTATTTTA 1200
QY 1201 GGAGAGAAACGCTGTTATTAATTAATCAATCCAGAAATGCTCACTGAGAAAGTTAAAGAA 1260
Db 1201 GGAGAGAAACGCTGTTATTAATTAATCAATCCAGAAATGCTCACTGAGAAAGTTAAAGAA 1260
QY 1261 ATTGAGATCGAATACAAATGAGACAGAGAGCTTCAAAACCGAACCGCTGGGGCTC 1320
Db 1261 ATTGAGATCGAATACAAATGAGACAGAGAGCTTCAAAACCGAACCGCTGGGGCTC 1320
QY 1321 CTGAGCAATGGAATGCGCTGCTTCCCTCTTCTAGGACCTCTAGCAGCTCTAATATTG 1380
Db 1321 CTGAGCAATGGAATGCGCTGCTTCCCTCTTCTAGGACCTCTAGCAGCTCTAATATTG 1380
QY 1381 TTACTCTCTTTGGACCTGATCTTAACTCTCTGTTAAGTTTGTCTCTCCAGAAAT 1440
Db 1381 TTACTCTCTTTGGACCTGATCTTAACTCTCTGTTAAGTTTGTCTCTCCAGAAAT 1440
QY 1441 GAAGCTGTAAAGCTACAGATGGTCTTTACAAATGGAACCCCA 1481

Db 1441 GAAGCTGTAAAGCTACAAATAGTCTTCAATGGAACCCCA 1481

RESULT 10
AC023366
LOCUS
DEFINITION

AC023366 172918 bp DNA linear HTG 02-MAR-2000
Homo sapiens clone RP11-15N10, WORKING DRAFT SEQUENCE, 33 unordered
pieces.

AC023366.2 GI:7143443

HTG: HTGS PHASE1: HTGS_DRAFT.

VERSION

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens, clone RP11-15N10

Unpublished

2 (bases 1 to 172918)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Beda, F., Boguslavsky, L.,

Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castie, A.,

Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,

DeArillano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,

Feneator, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,

Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Hoston, L.,

Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,

Klein, J., Landers, T., Largocque, K., Lehotzky, J., Levine, R.,

Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M.,

McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J.,

Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J.,

Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T. M.,

Peterson, K., Piere, N., Pisani, C., Pollara, V., Raymond, C.,

Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,

Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A.,

Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,

Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and

Zody, M.

Direct Submission

Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 2, 2000 this sequence version replaced gi:6970363.

All repeats were identified using RepeatMasker:

Smit, A.P.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L3492

Center clone name: 15 N.10

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 142569 bases at least Q40

Consensus quality: 156863 bases at least Q30

Consensus quality: 164107 bases at least Q20

Insert size: 178000; agarose-fp

Insert size: 169718; sum-of-contigs

Quality coverage: 2.9 in Q20 bases; agarose-fp

Quality coverage: 3.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 33 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

[illegible]

Db	1504	GAAGCTGTAAGCTACAAATGGTTCTTCTCAATAGGACCCCA	1544
RESULT 12	AY101589	10230 bp	DNA linear PRI 11-FEB-2004
LOCUS	AY101589		Gorilla gorilla isolate 1 endogenous retrovirus ERV-W, ERVW1
DEFINITION			locus, allele B, complete sequence.
ACCESSION	AY101589		
VERSION	AY101589.1	GI:37544413	
KEYWORDS			
SOURCE			Gorilla gorilla (gorilla)
ORGANISM			Gorilla gorilla
			Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
			Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
REFERENCE			1 (bases 1 to 10230)
AUTHORS			Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G.,
TITLE			Bonnaud, B., Lucotte, G., Duret, L., and Mandrand, B.
JOURNAL			The endogenous retroviral locus ERVW1 is a bona fide gene involved
PUBMED			in hominoid placental physiology
REFERENCE			Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
AUTHORS			14757826
TITLE			2 (bases 1 to 10230)
JOURNAL			Mallet, F., Bouton, O. and Oriol, G.
PUBMED			Direct Submission
REFERENCE			Submitted (06-MAY-2002) Retrovirology Department, UMR 2142
AUTHORS			CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée
TITLE			d'Italie, Lyon 69364 cedex 07, France
JOURNAL			Location/Qualifiers
FEATURES			1..10230
source			/organism="Gorilla gorilla"
			/mol_type="genomic DNA"
			/isolate="1"
			/isolation_source="PBMC of gorilla 1"
			/db_xref="taxon:9593"
			/cell_type="PBMC"
			/note="isolated by PCR; ERVW1 locus, allele B
			endogenous_virus: ERV-W"
			1..795
LTR			join(248..826,2894..9821)
mRNA			/note="putative mRNA transcript 1; gag-pol pseudogene"
			join(248..826,7595..9821)
mRNA			/note="putative mRNA transcript 2"
			join(248..826,9258..9821)
mRNA			/note="putative mRNA transcript 3; no coding region"
			7822..9438
CDS			/note="syncytin"
			/codon_start=1
			/product="envelope glycoprotein"
			/protein_id="AAM68168.1"
			/db_xref="GI:37544414"
			/translation="MALPHILLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLWRMRP
			GNIDAPSHRSFSGKTPTFTAHTHMPNCNSATLONHANTHYWTGKMINPSPGGILGV
			TVQWYFTFTGSDGGVQDQAREKHKEVISQLTRVHSSTSPYKGLDLSKHELTSTV
			HTLVSLNLTITGLHEVSAQNTNCIMCLPLDFRPVSIPTPEERNNSTEINTSV
			LVGLVSLNLEITHTSNLTCKVFSNTIDTNSQCIRWVTPPTQIVLPSPGFVCGTSA
			YRCLNGSSSMCFSLFVLPMTIYTBQDLNYYVSKRNKRVILFPIVIGAVLGAIG
			TGIGTTSQFYKLSQELNGDMERVADSLVTLQDLNSLAIVLQNRRLDLLTAE
			RGGTCLPLGEECCYVNSQSGIVTEKVEIRDRIQRRAEELRNTGPMGLLSOMMPWLP
			FLGLPAAIIILLIFGPCIFNLNLFVSSRIEAVLQMEPKWQSKTKIYRRLDRPASP
			RSDVNDIKCTPPEEISTAQPLLRPSAGSS"
			9497..10230
ORIGIN			
Query Match			90.6%; Score 1341.8; DB 9; Length 10230;
Best Local Similarity			94.1%; Pred. No. 0;
Matches 1394; Conservative			0; Mismatches 87; Indels 0; Gaps 0;
QY	1	ATGGCCCTCCCTATCATCTTTCTCTTTACTCTTTTACCTCTTCTCACT	60
Db	7822	ATGGCCCTCCCTATCATCTTTCTCTTTACTCTTTTACCTCTTCTCACT	7881

QY 61 GCACCCCTCCATGCTGCTGTAACAACAGTAGTCCCTTACCAGAGTTTCTATGAGA 120
Db 7882 GCACCCCTCCATGCTGCTGTAACAACAGTAGTCCCTTACCAGAGTTTCTATGAGA 7941
QY 121 ACGCGGCTTCCTGGAATATTTGATGCCCATATATAGGAGTTTATCTAAGGGAACCTCC 180
Db 7942 ATGCGGGTCTCTGGAATATTTGATGCCCATATATAGGAGTTTATCTAAGGGAACCTCC 8001
QY 181 ACTTCACTGCCACACCCCATATGCCCCGCAACTGCTATTAATCTGTCACCTCTTTGGATG 240
Db 8002 ACTTCACTGCCACACCCCATATGCCCCGCAACTGCTATTAATCTGTCACCTCTTTGGATG 8061
QY 241 CATGCAATACTCATTTATGGAAGGGAATATGATTAATCTAGTTGCTCTGGAGCACTT 300
Db 8062 CATGCAATACTCATTTATGGAAGGGAATATGATTAATCTAGTTGCTCTGGAGCACTT 8121
QY 301 GGAGCCACTGCTGTTGGACTTACTTCAACCATACAGTATGCTGATGGGGTGAAT 360
Db 8122 GGAGTCACTGCTGTTGGACTTACTTCAACCATACAGTATGCTGATGGGGTGAAT 8181
QY 361 CAAGGTCAGCAAGAGAAAAAAGTAAGGAAGCAATCTCCCAACTGACCCGGGACAT 420
Db 8182 CAAGATCAGGCAAGAGAAAAAAGTAAGGAAGCAATCTCCCAACTGACCCGGGACAT 8241
QY 421 AGCACCCCTAGCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAACCCCTCCGTACC 480
Db 8242 AGCACCTCTAGCCCTACAAAGGACTAGATCTCTCAAAACTACATGAACCCCTCCGTACC 8301
QY 481 CATACTCGCTGCTGAGCCTATTATTAACCAACCTCACTCGGCTCCATGAGGCTCAGCC 540
Db 8302 CATACTCGCTGCTGAGCCTATTATTAACCAACCTCACTCGGCTCCATGAGGCTCAGCC 8361
QY 541 CAAAACCTACTAATCTTGGATGTGCTCCCTCCCTGCATCTTCAAGCCATACATTTCAATC 600
Db 8362 CAAAACCTACTAATCTTGGATGTGCTCCCTCCCTGCATCTTCAAGCCATACATTTCAATC 8421
QY 601 CTGTTCTGAAACAATGGAAACAATTCAGACAGAAATAAACACCACTCCGTTTGTAGTA 660
Db 8422 CTGTTCTGAAACAATGGAAACAATTCAGACAGAAATAAACACCACTCCGTTTGTAGTA 8481
QY 661 GGACCTCTGTTTCCAACTCGGAATAACCACTCAACCTCAACCTCACTGCTGTAATAATTT 720
Db 8482 GGACCTCTGTTTCCAACTCGGAATAACCACTCAACCTCAACCTCACTGCTGTAATAATTT 8541
QY 721 AGCAATACTATAGACACAACAGCTCCCAATGATCAGGTGGGTAAACCTCCACACGA 780
Db 8542 AGCAATACTATAGACACAACAGCTCCCAATGATCAGGTGGGTAAACCTCCACACGA 8601
QY 781 ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGGTACCTCAGCCTATCATTTGTTG 840
Db 8602 ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGGTACCTCAGCCTATCATTTGTTG 8661
QY 841 AATGGCTCTTCAGAACTATGCTGCTCTCTCATTTCTAGTGGCCCTATGACCATCTAC 900
Db 8662 AATGGCTCTTCAGAACTATGCTGCTCTCTCATTTCTAGTGGCCCTATGACCATCTAC 8721
QY 901 ACTGAAACAAGATTATACAATCATGCTGCTACCTAAGCCCAACAACAAAGAGTACCCATT 960
Db 8722 ACTGAAACAAGATTATACAATCATGCTGCTACCTAAGCCCAACAACAAAGAGTACCCATT 8781
QY 961 CTTCTCTTTGTTATCAGACAGGAGTCTCAGGACAGACTAGTACTGTCATTTGGCAGPATC 1020
Db 8782 CTTCTCTTTGTTATAGGAGCAGGAGTCTAGGTGCACTAGTACTGTCATTTGGCGGTATC 8841
QY 1021 ACAACCTCTACTAGTTCTACTACAACATCTCAAGAAATAATGCTGATGGAACAG 1080
Db 8842 ACAACCTCTACTAGTTCTACTACAACATCTCAAGAAATAATGCTGATGGAACAG 8901
QY 1081 GTCACTGACTCCCTGGTCACTTCCAAAGATCAACTTAACCTCCCTAGCAGAGTAGTCTTT 1140
Db 8902 GTCCGCCACTCCCTGGTCACTTCCAAAGATCAACTTAACCTCCCTAGCAGAGTAGTCTTT 8961
QY 1141 CAAAATCGAAGAGCTTTAGACTTGTCTAAACCGCCAAAAGAGGGGGAACCTGTTTATTTTAA 1200

Db 8962 CAAAATCGAAGAGCTTTAGACTTGTCTAAACCGCTAAAAGAGGGGAACCTGTTTATTTTAA 9021
QY 1201 GGAGAAAGAACGCTGTTTATTTATGTTAATCAATCCAGAAATTTGCTACTGAGAAAGTTAAAGAA 1260
Db 9022 GGGGAAGAAATGCTGTTTATTTATGTTAATCAATCCGGAATGCTACTGAGAAAGTTAAAGAA 9081
QY 1261 ATTGAGATGCAATACATAGTAGAGAGAGCTTCAAAACACCCAGACCTGGGGCTC 1320
Db 9082 ATTGAGATGCAATACATAGTAGAGAGAGCTTCAAAACACCCAGACCTGGGGCTC 9141
QY 1321 CTCAGCAATGAGTGCCTCCCTCTTCCCTCTTAGGACCTCTAGCAGCTCTTAATATTG 1380
Db 9142 CTCAGCAATGAGTGCCTCCCTCTTCCCTCTTAGGACCTCTTAGCAGCTCTTAATATTG 9201
QY 1381 TTACTCTCTTTGGACCCCTGTATCTTTAACTCTCTTTGTTAAAGTTTGTCTCTCCAGAATT 1440
Db 9202 CTACTCTCTTTGGACCCCTGTATCTTTAACTCTCTTTGTTAACTTTGTCTCTCCAGAATC 9261
QY 1441 GAAGCTGTTAAAGCTACAGATGGTCTTTACAAATGGAACCCCA 1481
Db 9262 GAAGCTGTTAAAGCTACAAATGGAGCCCAAGATCGAGTCCAA 9302

RESULT 13
AC145964
LOCUS
DEFINITION Pan troglodytes BAC clone RP43-12F2 from 7, complete sequence.
ACCESSION AC145964
VERSION AC145964.3 GI:36016769
KEYWORDS HTG.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 184675)
AUTHORS Shahid, S., Cotton, M., Bielicki, L. and Meyer, R.
TITLE The sequence of Pan troglodytes BAC clone RP43-12F2
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 184675)
AUTHORS Sulston, J.E. and Wilson, R.
TITLE Sequencing of Pan troglodytes
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 184675)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 184675)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 184675)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (26-SEP-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 184675)
AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (29-OCT-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Sep 26, 2003 this sequence version replaced gi:35073486.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@wustl.edu
----- Summary Statistics
Center project name: C_PT012P02

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality ≥ 30); an attempt was made to resolve all sequencing problems, as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

MAPING INFORMATION: Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

SOURCE INFORMATION:
The RPCT-43 BAC Library has been constructed by Chung-Ii Shu. DNA was isolated from white blood cells obtained from a male chimpanzee (pan troglodytes, 'Clint', Yarkes #CQ471; birthdate: 6-6-80). The clone and detailed information can be obtained from ResGen (<http://www.resgen.com>) or Pier de Jong and co-workers at <http://www.bacpac.chori.org>.

NEIGHBORING SEQUENCE INFORMATION:

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.

FEATURES	source	Location/Qualifiers
		1. .184675 /organism="Pan troglodytes" /mol_type="genomic DNA" /db_xref="taxon:9598" /chromosome="7" /map="7" /clone="RP43-12P2" /clone_lib="RPC1-43" 115419..115500 /note="Sequence derived from one plasmid subclone." unsure 138694..138701 /note="Sequence derived from one plasmid subclone." unsure 140277..140388 /note="Sequence derived from PCR product of project DNA." misc_feature
ORIGIN		
Query Match		90.6%; Score 1341.8; DB 9; Length 184675;
Best Local Similarity		94.1%; Pred. No. 0;
Matches 1394;	Conservative	0; Mismatches 87; Indels 0; Gaps 0;
Qy	1	ATGCCCTCCCTTATCATCTTCTCTTTACTGTCTTACCCCTTTCGCCTCTCACT 60
Db	153623	ATGCCCTCCCTTATCATATTTCTCTTTACTGTCTTACCTCTTTCACCTCACT 153682
Qy	61	GCACCCCTCCATGTGCTGTACACCAAGTAGTCCCTTACCAGAGTTCTATGAAGA 120
Db	153693	GCACCCCTCCATGCTCGCTGTATGACCAAGTAGTCCCTTACCAGAGTTCTATGGAGA 153742
Qy	121	ACGGGGCTTCCTGGAAATATTGATGCCCCCATATAGGAGTTTATCTAAGGGAACCTCC 180
Db	153743	ATGCAGCGTCCCGGAAATATTGATGCCCATATAGAGAGTTTTTCTAAGGGAACCCCC 153802
Qy	181	ACCTTCACTGCCCCACACCCATATGCCCGCCAACTGCTATACTGCCACTCTTTGCATG 240
Db	153803	ACCTTCACTGCCACACCCATATGCCCGCCAACTGCTATCACTCTGCCACTCTTTGCATG 153862
Qy	241	CATGCAAAATCTCATTTATTGGACAGGAAAATGATTAATCTAGTTGTCCTCGAGAGACTT 300
Db	153863	CATGCAAAATCTCATTTATTGGACAGGAAAATGATTAATCTAGTTGTCCTCGAGAGACTT 153922
Qy	301	GGAGGCACATGCTGTGTGGACTTACTTCACCCATACCAAGTAGTCTGATGGGGGTGGAAATT 360

QY 1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
 Db 155063 GAAGCTGTAAAGCTACAAATGGAACCCCAAGATGAGTCCAA 155103

RESULT 14
 AY101586 10229 bp DNA linear PRI 11-FEB-2004
 LOCUS Pan troglodytes isolate 1 endogenous retrovirus ERV-W, ERVW1
 DEFINITION locus, allele A, complete sequence.
 ACCESSION AY101586
 VERSION AY101586.1 GI:37544407
 KEYWORDS Pan troglodytes (chimpanzee)
 SOURCE Pan troglodytes
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 REFERENCE 1 (bases 1 to 10229)
 AUTHORS Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G.,
 Bonnaud, B., Lucotte, G., Duret, L. and Mandrand, B.
 TITLE The endogenous retroviral locus ERVW1 is a bona fide gene involved
 in hominoid placental physiology
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
 PUBLISHED 14757826
 REFERENCE 2 (bases 1 to 10229)
 AUTHORS Mallet, F., Bouton, O. and Oriol, G.
 TITLE Direct Submission
 JOURNAL Submitted (06-MAY-2002) Retrovirology Department, UMR 2142
 CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée
 d'Italie, Lyon 69364 cedex 07, France

FEATURES
 source
 1..10229
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /isolate="1"
 /isolation_source="PBMC of chimpanzee 1"
 /db_xref="taxon:9598"
 /cell_type="PBMC"
 /notes="isolated by PCR; ERVW1 locus, allele A
 endogenous_virus: ERV-W"
 1..779
 join(248..810,2889..9820)
 /notes="putative mRNA transcript 1; gag-pol pseudogene"
 join(248..810,7592..9820)
 /product="envelope glycoprotein"
 /notes="putative mRNA transcript 2"
 join(248..810,9255..9820)
 /notes="putative mRNA transcript 3; no coding region"
 7819..9435
 /notes="syncytin"
 /codon_start=1
 /product="envelope glycoprotein"
 /protein_id="AAW68165.1"
 /db_xref="GI:37544408"
 /translation="MALPYHIFLFTVLPSFTLTAPPKCRMTSSSPYQBFRLMRORP
 GNIDAPSYRFSKGTPTFTATHPRNCHSATLCHMANTHYWTGKIILNPSCPGILGV
 TVCWYFTFTMGDDGGVQDAREKHVKEVISQLTRVHSAREPKGLDLSKLHETLR
 HTRLSFLNLTGLLEVSQAQNTWCNLCPLNRPYVSI PVPQWNFEINFTTSV
 LVGPLVNLBIHTSNLTCKVFSNTDTNQCIRWTPPTQIVCLPSGIFVCGTSA
 YRCLGSSSEKSLFLVPPWTIYTEQDLNYYVSKPRKRPVILPVPVIGAVLGALG
 TGLGTTSTFYTKLSQELNGDMERVADSLVTLQDLNSLAAVLQNRALDILLAE
 RGCTCLFLBEECCYYNQSGIVTEKVKSIHRIORRABELRNTGPMGLLSQWMPWILP
 FLAPLAAILLLFLGPCIENLVFVSSRIEAVKLQWEPKQSKTKIYRPLDRPASP
 RSDVNDIKGTPPEILLTAQPLLRPNAGSS"
 9494..10229

LTR
 ORIGIN
 Query Match 90.5%; Score 1340.2; DB 9; Length 10229;
 Best Local Similarity 94.1%; Pred. No. 0;
 Matches 1393; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
 QY 1 ATGGCCCTCCCTATCATCTTTCTTTACTGTTCTCTTACCCCTCTTTCACTCTCACT 7878
 61 GCACCCCTCCATGCTGCTGTCAACACAGTAGTCCCTTTACCAAGAGTTTCTATGAAGA 120
 7879 GCACCCCTCCATGCTGCTGTATGACACAGTAGTCCCTTTACCAAGAGTTTCTATGAGA 7938
 121 ACCGGCTTCTCGAATAATTGATGCCCATCATATAGGAGTTTATCTAAGGGGAACCTCC 180
 7939 ATGAGCGTCCCGGAATAATTGATGCCCATCATATAGGAGTTTCTTAAAGGAACCCCC 7998
 181 ACCTTCACCTGCCACACCCCATATGCCCCCGCAACTGCTATATACTCTGCACTCTTTGCATG 240
 7999 ACCTTCACCTGCCACACCCCATATGCCCCCGCAACTGCTATACCTCTGCACTCTTTGCATG 8058
 241 CATGCAATACTCATTTATGGAAGGAAATGATTAATCTCTAGTTCTCTGGAGGACTT 300
 8059 CATGCAATACTCATTTATGGAAGGAAATGATTAATCTCTAGTTCTCTGGAGGACTT 8118
 301 GGAGCCACTGCTGTTGGACTTACTTCAACCACACAGTATGCTGATGGGGTGGAAAT 360
 8119 GGAGTCACTGCTGTTGGACTTACTTCAACCACACAGTATGCTGATGGGGTGGAGTT 8178
 361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCCGGGACAT 420
 8179 CAAGATCAGGCAAGAGAAAAACAAGTAAAGGAAGTAAATCTCCCAACTCACCCGGGTACAT 8238
 421 AGCACCCCTAGCCCTCAAAAGGACTAGTTCTCTCAAAACTCATATGAACCCCTCCGTACC 480
 8239 AGCGCCTTAGGCCCTACAAAGGACTAGATCTCTCAAAACTCATATGAACCCCTCCGTACC 8298
 481 CATACTGCTGCTGAGCCCTATTTAATACCAACCTCACTCGGCTCCATGAGGTCTACGC 540
 8299 CATACTGCTGCTGAGCCCTATTTAATACCAACCTCACTCGGCTCCATGAGGTCTCGGCC 8358
 541 CAAAACCTTACTTAACCTGTTGGATGCTCTCCCTCTGCACTTTCAGGCCATACATTTCAATC 600
 8359 CAAAACCTTACTTAACCTGTTGGATGCTCTCCCTCTGCACTTTCAGGCCATATGTTTCAATC 8418
 601 CTGTTCTCGAAACAATGGAAACAACCTTCAAGACAGAAATAAACCACCTTCCGTTTTAGTA 660
 8419 CTTGTACCTGAAACAATGGAAACAACCTTCAAGACAGAAATAAACCACCTTCCGTTTTAGTA 8478
 661 GGACCTCTGTTTCCCAATCTGGAATAAACCATACCTCAACCTCACCTCTGTATAAATTT 720
 8479 GGACCTCTGTTTCCCAATCTGGAATAAACCATACCTCAACCTCACCTCTGTATAAATTT 8538
 721 AGCAATACTATAGACACAACACCTCCCAATGATCAGGTGGGTAAACCTTCCACACGA 780
 8539 AGCAATACTATAGACACAACACCTCCCAATGATCAGGTGGGTAAACCTTCCACACGA 8598
 781 ATAGTCTGCTACCCCTCAGGAATATTTTGTCTGTGGTACCTCAGCTATCATTTGTTG 840
 8599 ATAGTCTGCTACCCCTCAGGAATATTTTGTCTGTGGTACCTCAGCTATCATTTGTTG 8658
 841 AATGGCTCTTCAGAATCTATGCTCTCTCTCATTTCTAGTGGCCCCCTATGACCATCTAC 900
 8659 AATGGCTCTTCAGAATCTATGCTCTCTCTCATTTCTAGTGGCCCCCTATGACCATCTAC 8718
 901 ACTGAAACAAGATTATACAAATCATGCTGTAACCTTAAGCCCCCAACAAGAGTACCCATT 960
 8719 ACTGAAACAAGATTATACAAATCATGCTGTAACCTTAAGCCCCCAACAAGAGTACCCATT 8778
 961 CTTCTCTTTTGTATCAGAGCAGGAGTGTAGGAGACTAGGTACTTGGCATTGGCAGTATC 1020
 8779 CTTCTCTTTTGTATCAGAGCAGGAGTGTAGGAGACTAGGTACTTGGCATTGGCAGTATC 8838
 1021 ACAACCTCTACTCAGTTCTTACTACAACTATCTCAAGAAATAAATGGTGTGACATGGAAACAG 1080
 8839 ACAACCTCTACTCAGTTCTTACTACAACTATCTCAAGAAATAAATGGTGTGACATGGAAACAG 8898
 1081 GTCACTGACCTCCCTGCTGACCTTGCAGATCACTTAACTCCCTAGCAGAGTAGTCTCTT 1140
 8899 GTCCGCGACTCCCTGCTGACCTTGCAGATCACTTAACTCCCTAGCAGAGTAGTCTCTT 8958

```
1141 CAAAATCGAAGCTTTAGACTTGTCTAAACCGCCAAAGAGGGGACCTGTTTATTTT 1200
|
|
|
8959 CAAAATCGAAGAGCTTTAGACTTGTCTAAACCGCTGAAAGAGGGGAAACCTGTTTATTTT 9018
|
|
|
1201 GGAAGAGAGCGCTGTTTATTATGTTAATCAATCAGAAATTTGTCACCTGAGAAAGTTAAAGAA 1260
|
|
|
9019 GGGGAAGAATGCTGTTTATTATGTTAATCAATCGGGAATCGTCACTGAGAAAGTTAAAGAA 9078
|
|
|
1261 ATTTCGAGATCGAATACAATGTTAGAGCAGAGAGCTTCAAACACCGAAGCTGGGGCCTC 1320
|
|
|
9079 ATTTCGAGATCGAATACAATGTTAGAGCAGAGAGCTTCAAACACCGAAGCTGGGGCCTC 9138
|
|
|
1321 CTAGCCCAATGATGCGCTGGGTTCTCCCTTTTAGGACCTCTAGACGCTCTAATATG 1380
|
|
|
9139 CTAGCCCAATGATGCGCTGGGTTCTCCCTTTTAGGACCTCTAGACGCTCTAATATG 9198
|
|
|
1381 TTAATCTCTTTGACCGCTGATCTTAACTCTCTGTTAAAGTTTGTCTTCCAGAAAT 1440
|
|
|
9199 CTACTCTCTTTGACCGCTGATCTTAACTCTCTGTTAAAGTTTGTCTTCCAGAAAT 9258
|
|
|
1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
|
|
|
9259 GAAGCTGTAAAGCTACAAATGGAACCCCAAGATGAGTCCAA 9299
|
|
|
RESULT 15
AY101587
LOCUS
DEFINITION Pan troglodytes isolate 1 endogenous retrovirus ERV-W, ERVW1
locus, allele B, complete sequence.
ACCESSION AY101587
VERSION
KEYWORDS
SOURCE
ORGANISM Pan troglodytes (chimpanzee)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
AUTHORS Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
TITLE The endogenous retroviral locus ERVW1 is a bona fide gene involved
in hominoid placental physiology
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
PUBMED 14757826
REFERENCE
AUTHORS Mallet,F., Bouton,O. and Oriol,G.
TITLE Direct Submission
JOURNAL Submitted (06-MAY-2002) Retrovirology Department, UMR 2142
CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée
d'Italie, Lyon 69364 cedex 07, France
FEATURES
source
1. .10229
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/isolate="1"
/isolation_source="PBMC of chimpanzee 1"
/db_xref="taxon:9598"
/cell_type="PBMC"
/notes="isolated by PCR; ERVW1 locus, allele B
endogenous_virus: ERV-W"
1. .779
join(248. .810,2889. .9820)
/notes="putative mRNA transcript 1; gag-pol pseudogene"
join(248. .810,7592. .9820)
/product="envelope glycoprotein"
/notes="putative mRNA transcript 2"
join(248. .810,9255. .9820)
/notes="putative mRNA transcript 3; no coding region"
7819. .9435
/notes="syncytin"
/codon_start=1
/product="envelope glycoprotein"
```

```
/protein_id="AAM68166.1"
/db_xref="gi:37544410"
/translation="MALPFIHIFLFTVLPSFTLAPPRCMTSSSPVQEFWMWRP
GNIDAPSVRSKGTPTTAHTMPKNSHSATLPMHANTHWTKGLINPSPGSLGV
TVQVTFHTGMSDGGVQDOAREKHVKEVISQLTRVHSASRPYKGLDLSKLHELTR
HTRVLSPNLTITLGLHEVSAQNPNCIMCLPLNRPYVSIIVPQWNNFSTEINTSV
LVGLNSESSENCFLSFLVPPMTIYTEQDLNYYVSKPKRVPILPFFVIGAGVLCALG
YRCPLNGSSENCFLSFLVPPMTIYTEQDLNYYVSKPKRVPILPFFVIGAGVLCALG
TGIGITTTSTQFYKLSQELNGDMRVADSLVTIQDQNLNLAALVQNRALDLTAE
RGCTCLFLGSECCYVNGSGIVTEKVEIRDRIOQRAEELNTGFWGLLSQWMPILP
PLGLAAAILLILLPGPCIENLLNVPVSSRIEAVKLQMEPKMQSKTKYRRLPLDRDASP
RSDVNDIKGTPEPEILTAQPLLRPNASGS"
9494. .10229
LTR
ORIGIN
Query Match 90.5%; Score 1340.2; DB 9; Length 10229;
Best Local Similarity 94.1%; Pred. No. 0;
Matches 1393; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 1 ATGGCCCTCCCTTATCATCTTTTCTTTTACTGTCTTACCCCTTTCCTCTCACT 60
DB 7819 ATGGCCCTCCCTTATCATCTTTTCTTTTACTGTCTTACCCCTTTCCTCTCACT 7878
QY 61 GCACCCCTCCATGCTGCTGTACACACCGTAGTCCCTTACCAAGAGTTTCTATGAAGA 120
DB 7879 GCACCCCTCCATGCTGCTGTATGACCATGCTCCCTTACCAAGAGTTTCTATGAAGA 7938
QY 121 ACGCGCTTCTCGGAAATATTGATGCCCATCATATAGGAGTTTATCTAAGGGAATCTCC 180
DB 7939 ATGCAGCTGCTCGGAAATATTGATGCCCATCATATAGGAGTTTCTAAGGGAATCTCC 7998
QY 181 ACCTTCACCTGCCCCACACCATATGCCCCCACTGCTATACTCTGCCACTCTTTGCATG 240
DB 7999 ACCTTCACCTGCCCCACACCATATGCCCCCACTGCTATACTCTGCCACTCTTTGCATG 8058
QY 241 CATGCAATACCTACTTATTGGAGAGGAAATGATTAAATCTAGTTGTCTCTGGAGACTT 300
DB 8059 CATGCAATACCTACTTATTGGAGAGGAAATGATTAAATCTAGTTGTCTCTGGAGACTT 8118
QY 301 GGAGCCACTGCTGTGTGGACTTACTTACCCCATACCAATGATGCTATGCTGATGGGGTGAATT 360
DB 8119 GGAGTCACTGCTGTGTGGACTTACTTACCCCATACCAATGATGCTGATGGGGTGAATT 8178
QY 361 CAAGTCAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 8179 CAAGATCAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8238
QY 421 AGCACCCCTAGCCCTACAAAGAGCTAGTTCTCTCAAAATACATGAAACCCCTCCGTACC 480
DB 8239 AGCGCTCTAGGCCCTACAAAGAGCTAGATCTCTCAAAATACATGAAACCCCTCCGTACC 8298
QY 481 CATACTCGCTGTGAGCCCTATTATATACACCCCTCACTCGGCTCCATGAGGTCTCAGCC 540
DB 8299 CATACTCGCTGTGAGCCCTATTATATACACCCCTCACTCGGCTCCATGAGGTCTCAGCC 8358
QY 541 CAAAACCTACTACTGTTGGATGCTCCCTCCCTGACCTTCAGGCCATACATTTCAATC 600
DB 8359 CAAAACCTACTACTGTTGGATGCTCCCTCCCTGACCTTCAGGCCATACATTTCAATC 8418
QY 601 CTGTTCTCTGAACAATGGAAACAACTTTCAGACAGAAATAAACACCACTTCGTTTTAGTA 660
DB 8419 CTGTACTCTGAACAATGGAAACAACTTTCAGACAGAAATAAACACCACTTCGTTTTAGTA 8478
QY 661 GGACCTCTGTTTCCAATCTGGAATAAACCACTACTCAACCTCAACCTCACTGTTGTAATAATT 720
DB 8479 GGACCTCTGTTTCCAATCTGGAATAAACCACTACTCAACCTCAACCTCACTGTTGTAATAATT 8538
QY 721 AGCAATACTATAGACACAAACAGCTCCCAATGATCAGGTGGGTAAACCTCCACACGA 780
DB 8539 AGCAATACTATAGACACAAACAGCTCCCAATGATCAGGTGGGTAAACCTCCACACGA 8598
QY 781 ATAGTCTGCCTACCTCAGGAATATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 840
|
|
|
```


Db	8302	CATACCTCGCTGGTAGCCCTATTAAATACCAACCCCTCACTGGGCTCCATGAGGTCTCGGCC	8361
Qy	541	CAAAAACCTACTAACCTGTGGATGTGCTCCCTGCATCTTCAGGCCATACATTTCAATC	600
Db	8362	CAAAAACCTACTAACCTGTGGATGTGCTCCCTGCATCTTCAGGCCATACATTTCAATC	8421
Qy	601	CTGTGTTCCCTGAAACAATGGAAACAATTCAGACACAGAAATAAAACACATTCGGTTTTAGTA	660
Db	8422	CTGTGTTACCTGAAGATGGAAACAATTCAGACACAGAAATAAAACACATTCGGTTTTAGTA	8481
Qy	661	GGACCTCTGTTTCCAACTCTGGAATAAACCCATACTCTCAAACTCACCTGTGTAAATTT	720
Db	8482	GGACCTCTGTTTCCAACTCTGGAATAAACCCATACTCTCAAACTCACCTGTGTAAATTT	8541
Qy	721	AGCAATFACTATAGACACAACACAGCTCCCAATGCATCAGGTGGGTAAACCTCCACACGA	780
Db	8542	AGCAATFACTATAGACACAACACAGCTCCCAATGCATCAGGTGGGTAAACCTCCACACGA	8601
Qy	781	ATAGTCTGCGCTACCCCTCAGGAATATTTTGTCTGTGGTACCTCAGCCTATCATTTGTTG	840
Db	8602	ATAGTCTGCGCTACCCCTCAGGAATATTTTGTCTGTGGTACCTCAGCCTATCATTTGTTG	8661
Qy	841	AATGGCTCTTCAGAACTATGTGCTCTCTCAATCTTAGTGCCCTATGACCACTTAC	900
Db	8662	AATGGCTCTTCAGAACTATGTGCTCTCTCAATCTTAGTGCCCTATGACCACTTAC	8721
Qy	901	ACTCAACAAGATTTATACAATCATGTGTTACCTTAAGCCCAACAACAAAGATGCCATT	960
Db	8722	ACTGAAACAAGATTTATACAATCATGTGTTACCTTAAGCCCAACAACAAAGATGCCATT	8781
Qy	961	CTTCCTTTTGTATACAGACGAGGAGTCTAGGCAGACTAGGTACTGGCATTCGGCAGTATC	1020
Db	8782	CTTCCTTTTGTATAGGAGCGGGAGTCTAGGTGCTAGGTACTGGCATTCGGCAGTATC	8841
Qy	1021	ACAACCTCTACTCAGTTTCTACTACAAACTATCTCAAGAAATAAATGGTGAATGGAACAG	1080
Db	8842	ACAACCTCTACTCAGTTTCTACTACAAACTATCTCAAGAAATAAATGGTGAATGGAACAG	8901
Qy	1081	GTCACTGACTCCCTGGTCACTCTGCAAGATCACTTAACCTCCCTAGCAGCGTAGTCCTT	1140
Db	8902	GTCACTGACTCCCTGGTCACTCTGCAAGATCACTTAACCTCCCTAGCAGCGTAGTCCTT	8961
Qy	1141	CAAAATCGAAGAGCTTTAGACTGTCTAACCGCCAAAAGAGGGGAACTGTTTATTTTA	1200
Db	8962	CAAAATCGAAGAGCTTTAGACTGTCTAACCGCCAAAAGAGGGGAACTGTTTATTTTA	9021
Qy	1201	GGAGAAGAACGCTGTATTATGTATTAATCAATCCGAATTTGCTACTGAGAAGTTAAAGAA	1260
Db	9022	GGAGAAGAACGCTGTATTATGTATTAATCAATCCGAATTTGCTACTGAGAAGTTAAAGAA	9081
Qy	1261	ATTTCGAGATCGAATACAATGTAGACAGAGGAGCTTCAAAACACCGAACCGCTGGGCGCT	1320
Db	9082	ATTTCGAGATCGAATACAATGTAGACAGAGGAGCTTCAAAACACCGAACCGCTGGGCGCT	9141
Qy	1321	CTCAGCCCAATGGATGCCCTGGGTTCTCCCTTTCTTAAGACCTCTAGCAGCTCTAATTTG	1380
Db	9142	CTCAGCCCAATGGATGCCCTGGGTTCTCCCTTTCTTAGGACCTCTCTGAGCTATAATTTG	9201
Qy	1381	TTACTCTCTTTGGACCTGTATCTTTAACTCTCTTTGTTAAAGTTTGTCTCTTCAGAAATT	1440
Db	9202	CTACTCTCTTTGGACCTGTATCTTTAACTCTCTTTGTTAACTTGTCTCTTCAGAAATT	9261
Qy	1441	GAAGCTGTAAAGCTACAGATGTTCTTACAATGGAACCCCA	1481
Db	9262	GAAGCTGTAAAGCTACAGATGTTCTTACAATGGAACCCCA	9302

RESULT 17

AY101593

LOCUS

DEFINITION

10246 bp

DNA

linear

PRI 11-FEB-2004

Hylobates pileatus isolate 1

endogenous retrovirus ERV-W, ERVME1

locus, allele B, complete sequence.

181 ACCTTCACTGCGCACCCCATATGCGCGCAACGTCTATTAACCTTGCACCTCTTGCATG 240
7986 ACCTTCACTGCGCACCCCATATATGCGCGCAACGTCTATTAACCTTGCACCTCTTGCATG 8045
241 CATGCAAAATCTCATTTATGAGACAGGAAAATGATTAATCCTAGTTGTCTCGAGGACATT 300
8046 CATGCAAAATCTCAGTATTGACAGGAAAATGATTAATCCTAGTTGTCTCGAGGACATT 8105
301 GGAGCCACTCTCTGTTGGACTTACTTCAACCATACCAAGTATGCTCGATGGGGTGGAAATT 360
8106 GGAGTCACTCTCTGTTGGACTTACTTCAACCATACCAAGTATGCTCGATGGGGTGGAGTT 8165
361 CAAGGTCAGGCAAGAGAAAACAAAGTAAAGGAGCAATCTCCCAACTGACCCGGGACAT 420
8166 CAAGATCAGGCAAGAGAAAACAAAGTAAAGGAGTAACTTCCCAACTGACCCAGGTACAT 8225
421 AGCACCCCTAGCCCTACAAAGGAGTAGTTCTCTCAAAAATACATGAAACCCCTCCGTACC 480
8226 AGCACCTCTAGCCCTACAAAGGAGTAGATCTCTCAAAATCATGAAACCCCTCCGTACC 8285
481 CATACTCGCTGTGAGCCTATTAAATACACCCCTCACTCGGCTCCATGAGGTCTCAGCC 540
8286 CACACTCGCTGTGAGCCTATTAAATACACCACTCACTGGGCTCCATGAGGTCTCGGCC 8345
541 CAAAACCTACTAACTGTTGGATGCGCTCCCTCGACTTCTAGGCCATACATTTCAATC 600
8346 CAAAACCTACTAACTGTTGGATGCGCTCCCTCGACTTCTAGGCCATATGTTTCAATC 8405
601 CTTGTTCTCGAACAATGGAACAACCTTCCAGACAGAAATAAACACCACTTCCGTTTGTAGTA 660
8406 CTTGTACTGAAACANTGGAACAACCTTCCAGACAGAAATAAACACCACTTCCGTTTGTAGTA 8465
661 GGACCTCTGTTTCCAACTCGAAATAACCATACCTCCTCAACCTCACTGTGTGAAAATTT 720
8466 GGACCTCTGTTTCCAACTCGAAATAACCATACCTTCAACCTCACTGTGTGAAAATTT 8525
721 AGCAATACTATAGACACACCACTCCCAATGATCAGTGGGTAAACCTCCCAACGA 780
8526 AGCAATACTATAGACACACCACTCCCAATGATCAGTGGGTAAACCTCCCAACGA 8585
781 ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGGTACCTCAGCTATCATCTGTTTG 840
8586 ATATCTGCTACCTCAGGAATATTTTGTCTGTGGTACCTCAGCTATCATCTGTTTG 8645
841 AATGGCTCTTCAGAATCTATGTCCTTCTCTCAATCTTATGTGCCCCCTATGACCATCTAC 900
8646 AATGGCTCTTCAGAATCTATGTCCTTCTCTCAATCTTATGTGCCCCCTATGACCATCTAC 8705
901 ACTGAACAAGATTATACATCATGTCGTACCTAACGCCCAACAAAGAGTACCATT 960
8706 ACTGAACAAGATTATACATCATGTCGTACCTAACGCCCGCAACAAAGAGTACCATT 8765
961 CTTCTCTTTGTTATCAGACAGAGTGTAGGACAGACTAGGTACTGSCATTGGCAGTATC 1020
8766 CTTCTCTTTGTTATGGAGCAGAGTGTAGGTGCACTAGGTACTGGCATTTGGCAGTATC 8825
1021 ACAACCTCTACTCAGTTCTACTACAAATCTCTCAAGAAATAAATGTTGACATGGAACAG 1080
8826 ACAACCTCTACTCAGTTCTACTACAAATCTCTCGAAGAAATAAATGTTGACATGGAACG 8885
1081 GTCATGACTCCCTGTCACCTTGGCAAGATCAACTTAATCTCCCTAGCAGCAGTAGTCTTT 1140
8886 GTCGCGACTCCCTGTCACCTTGGCAAGATCAACTTAATCTCCCTAGCAGCAGTAGTCTTT 8945
1141 CAAAATCGAGAGCTTTAGACTTGTCTAAACGCAAGAGGGGGAACCTGTTTATTTTAA 1200
8946 CAAAATCGAGAGCTTTAGACTTGTCTAAACGCGCTGAAAGAGGGGGAACCTGTTTATTTTAA 9005
1201 GGAGAGAAACCGCTGTTTATTTATGTTAATCAATCCAGAAATGTTCTACTGAGAAAGTTAAAGAA 1260
9006 GGAGAGAAATGCTGTTTATTTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 9065
1261 ATTCGAGATCGAATAACAATGTAGACAGAGGAGCTTTCAAAACACCGAACCGCTGGGCGCTC 1320

1066 ATTTCGAGATCGAATACAAGTAGACAGAGAGCTTCGAAACATTTGGACCTCTGGGCGCTC 9125
1321 CTGAGCAATGAGTGCCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATAATTG 1380
9126 CTGAGCAATGAGTGCCTGGATTCCTCCCTTCTTAGGACCTCTAGCAGCTATAAATTG 9185
1381 TTAATCTCTTTGGAGACCTGTATCTTTAACTCCTCTTGTAAAGTTTGTCTTCCAGAAAT 1440
9186 CTGCTCTCTTTGGAGACCTGTATCTTTAACTCCTCTTGTAAAGTTTGTCTTCCAGAAAT 9245
1441 GAAGCTGTAAGCTACAGATGCTTTACAAATGGAACCCCA 1481
9246 GAAGCTATTAACATAAAATGGAGCCCAAGATGGATGCCAA 9286
RESULT 18
AY101592 10248 bp DNA linear PRI 11-FEB-2004
LOCUS Hylobates pileatus isolate 1 endogenous retrovirus ERV-W, ERVWE1
DEFINITION locus, allele A, complete sequence.
ACCESSION AY101592
VERSION AY101592.1 GI:37544419
KEYWORDS Hylobates pileatus (pileated gibbon)
SOURCE Hylobates pileatus
ORGANISM Hylobates pileatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.
REFERENCE 1 (bases 1 to 10248)
AUTHORS Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
The endogenous retroviral locus ERVWE1 is a bona fide gene involved
in hominoid placental physiology
Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
14757826
2 (bases 1 to 10248)
Mallet,F., Bouton,O. and Oriol,G.
Direct Submission
Submitted (06-MAY-2002) Retrovirology Department, UMR 2142
CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée
d'Italie, Lyon 69364 cedex 07, France
FEATURES
source
1. .10248
/organism="Hylobates pileatus"
/mol_type="genomic DNA"
/isolate="1"
/isolation_source="PBMC of gibbon 1"
/db_xref="Taxon:9589"
/cell_type="PBMC"
/note="isolated by PCR; ERVWE1 locus, allele A
endogenous_virus: ERV-W"
1. .773
join(248..804,2997..9809)
/note="putative mRNA transcript 1; gag-pol pseudogene"
join(248..804,7581..9809)
/product="envelope glycoprotein"
/note="putative mRNA transcript 2"
join(248..804,9244..9809)
/note="putative mRNA transcript 3; no coding region"
7808..9424
/note="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAM68171.1"
/db_xref="GI:37544420"
/translation="MALPHIFLFTVLLPSLTAPPCCRCMTSSSPYQEFILWRTORP
GNIDPLVRSFGSKGPTFTAHYMPRTCYNSATLCHANTQYTMGNINPSPGGIGV
TVCWTFYTHGSDGGVQDQAREKHVKEVISQLTQVHSTSSPYKGLDLISKLHETLRT
HPLVSLFNTLTLGLHESAOPTNCWMLPLDFRYPVSIPIVPEOWNNFSLEINTTSV
LGPLVSNLEITHTLNLCVKFSNTTDTTNSOCIRWVTPTRI FCLPSGIFVCGTSA
YCLNGSSESMCFSLFPLPDPWITYTEODLYNTVSKPRKRPVLPFVNGAVLGAIG
TGIGSITTSQFYKLSELNGDMERVA DSLVTLQDLNSLAALVQLNRRALDILLTAE
RGGTCLFLGECCYYNQSQSIIVTEKYKEIRDIRRAEELRNIGWGLLSQWMPWILP

ORIGIN		FLGPLAAIIILLIFGCIENILNVFVSSRIEAIKLOMBPKMESKTKNVRSLDWPASP
LTR		RSVDNDIKGIPPEIEISTAOLLRPNAGSS"
ORIGIN		9483..10248
Query Match		90.4%; Score 1338.6; DB 9; Length 10248;
Best Local Similarity		94.0%; Pred. No. 0;
Matches 1392; Conservative		0; Mismatches 89; Indels 0; Gaps 0;
QY	1	ATGCGCCCTCCCTTATCATATCTTTCTTTTACHTGTTCTCTTACCCCTTTGCTCTCACT 60
DB	7808	ATGCGCCCTCCCTTATCATATCTTTCTTTTACHTGTTCTCTTACCCCTTTGCTCTCACT 7867
QY	61	GCACCCCTCCATGCTGTGACACACGAGTAGTCCCTTACCAAGAGTTTCTATGAAGA 120
DB	7868	GCACCCCTCCATGCTGTATGACAGTAGTCCCTTACCAAGAGTTTCTATGAGA 7927
QY	121	ACGGGCTTCTCGAATAATTTGATGCCCATATATAGGAGTTTATCTAAGGGAACCTCC 180
DB	7928	ACGAGCGTCTCGAATAATTTGATGCCCATATATAGGAGTTTATCTAAGGGAAGCCCC 7987
QY	181	ACCTTCACTGCCACACCCATATGCCCCGCAACTGCTATTAACCTCTGCACTCTTTCATG 240
DB	7988	ACCTTCACTGCCACACCTATATGCCCCGCAACTGCTATTAACCTCTGCACTCTTTCATG 8047
QY	241	CATGCAATACCTATTTATGACAGGGAATAATGATTAATCCTAGTTGTTCTCTGAGGACTT 300
DB	8048	CATGCAATACCTAGTATTTGACAGGGAATAATGATTAATCCTAGTTGTTCTCTGAGGACTT 8107
QY	301	GGAGCCACTGCTCTGGACTTACTTCAACCACATACAGTATGCTGTATGGGGTGGAAAT 360
DB	8108	GGAGTCACTGCTCTGGACTTACTTCAACCACATACAGTATGCTGTATGGGGTGGAGTT 8167
QY	361	CAAGGTGAGCAAGAGAAAACAAGTAAAGGAACATCTCCCACTGACGCCGGGACAT 420
DB	8168	CAAGATCAGCAAGAGAAAACAATGTAAGAAGTAATCTCCCACTGACCCAGGTATCAT 8227
QY	421	AGCACCTTAGCCCTACAGAGGACTGTTCTCTCAAACTACATGAACCTCCCTCGTACC 480
DB	8228	AGCACCTTAGCCCTACAGAGGACTGATGTTCTCAAACTACATGAACCTCCCTCGTACC 8287
QY	481	CATACCTCGCTGGTGAGCTTATTAATACCACTCTACCTGCTGCTGATGAGGTCTAGCC 540
DB	8288	CACACTCGCTGTTAAGCTTATTAATACCACTCTACCTGCTGCTGATGAGGTCTAGCC 8347
QY	541	CAAAACCTTACTATGTTGATGCTCCCTGCACTTTCAGGCCATACATTTCAATC 600
DB	8348	CAAAACCTTACTATGTTGATGCTCCCTGCACTTTCAGGCCATATGTTTCAATC 8407
QY	601	CTGTGTTCTGACAAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTATGTA 660
DB	8408	CTGTGTTCTGACAAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTATGTA 8467
QY	661	GGACCTCTGTTTCCAAATCTGGAATAACCCATACCTCAAACTCACCTGTTGTAATTT 720
DB	8468	GGACCTCTGTTTCCAAATCTGGAATAACCCATACCTCAAACTCACCTGTTGTAATTT 8527
QY	721	AGCAATACATAGACACACACGCTCCCAATGATCATGCTGGTGAACACCTCCACACGA 780
DB	8528	AGCAATACATAGACACACACGCTCCCAATGATCATGCTGGTGAACACCTCCACACGA 8587
QY	781	ATAGTCTGCTACCTCCAGAAATATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 840
DB	8588	ATATCTGCTACCTCCAGAAATATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 8647
QY	841	AATGGCTCTCAGAACTATGTTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
DB	8648	AATGGCTCTCAGAACTATGTTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8707
QY	901	ACTGACACAGATTTATACAACTATGCTGCTACCTAGACCCCAACAAAGAGTACCCATT 960
DB	8708	ACTGACACAGATTTATACAACTATGCTGCTATCTAGCCCCCGACAAAGAGTACCCATT 8767

QY	961	CTTCTCTTTGTTATCAGAGCAGGAGTGCTAGGACAGACTAGGTACTTGCGATTGGCAGTATC 1020
DB	8768	CTTCTCTTTGTTATGGGACAGGAGTGCTAGGTCAGTACTAGGTACTTGCGATTGGCAGTATC 8827
QY	1021	ACAACCTTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACATGAACAG 1080
DB	8828	ACAACCTTCTACTCAGTTCTACTACAAACTATCTCAGAACTAAATGGTGACATGAACAG 8887
QY	1081	GTCACCTGACTCCCTGGTCACCTTGCAGATCAACTTAATCTCCCTAGCAGCAGTAGTCTTT 1140
DB	8888	GTCGCGACTCCCTGGTCACCTTGCAGATCAACTTAATCTCCCTAGCAGCAGTAGTCTTT 8947
QY	1141	CAAAATCGAAGAGCTTTAGACTTGTCAACCCCAAGAGGGGGAACCTGTTTATTTT 1200
DB	8948	CAAAATCGAAGAGCTTTAGACTTGTCAACCCGCTGAAGAGGGGGAACCTGTTTATTTT 9007
QY	1201	GGAGAAAGACGCTGTTATTTATTAATCAATCCAGAAATGTCACCTGAGAAAGTTAAAGAA 1260
DB	9008	GGGGAAGAATGCTGTTATTAATCAATCCGGAATCGTCACCTGAGAAAGTTAAAGAA 9067
QY	1261	ATTCGAGATCGAATACAATGTAGACAGAGGAGTTCAAAACACCGAACGCTGGGGCTC 1320
DB	9068	ATTCGAGATCGAATACAACGTAGACAGAGGAGTTCGAAAACATTTGGACCTGGGGCTC 9127
QY	1321	CTCAGCAATGAGTCCCTGGTCTCCCTCTTTAGGACCTCTAGCAGCTCTAATATTG 1380
DB	9128	CTCAGCAATGAGTCCCTGGATCTCCCTCTTTAGGACCTCTAGCAGCTCTAATATTG 9187
QY	1381	TTACTCTCTTTGGACCTCTGATCTTTAACTCTTTTAACTCTTTTAACTCTTTTAACTCT 1440
DB	9188	CTGCTCTCTTTGGACCTGATCTTTAACTCTTTTAACTCTTTTAACTCTTTTAACTCT 9247
QY	1441	GAAGCTGTAAAGCTCAGATGCTTACAAATGGAACCCCA 1481
DB	9248	GAAGCTGTAAAGCTCAGATGGAAGCCCAAGATGGAGTCCAA 9288
RESULT 19		
AF520477S2		2694 bp DNA linear PRI 11-FEB-2004
LOCUS		Homo sapiens individual 132 allele A, envelope glycoprotein gene,
DEFINITION		complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION		AF520478
VERSION		AF520478.1 GI:33410924
KEYWORDS		2 of 2
SEGMENT		
SOURCE		Homo sapiens (human)
ORGANISM		Homo sapiens
REFERENCE		Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE		Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G.,
JOURNAL		Bonnaud, B., Lucotte, G., Duret, L. and Mandrand, B. A bona fide gene involved
PUBMED		The endogenous retroviral locus ERVWE1 is a placental physiology
AUTHORS		in hominoid placental physiology
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
REFERENCE		14757826
AUTHORS		2 (bases 1 to 2694)
TITLE		Mallet, F., Bouton, O. and Oriol, G.
JOURNAL		Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
PUBMED		CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée
AUTHORS		d'Italie, Lyon 69364 cedex 07, France
JOURNAL		Location/Qualifiers
FEATURES		1. .2694
source		/organism="Homo sapiens"
		/mol_type="genomic DNA"
		/isolate="individual 132 allele A"
		/db_xref="taxon:9606"
		/chromosome="7"
		/map="7q21-q22"
		/sex="female"
		/cell_type="PBMC's"

```
/note="Caucasian
endogenous_virus: HERV-W"
misc_feature
43..57
/note="splice acceptor site"
CDS
284..1900
/note="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAQ17561.1"
/db_xref="GI:33410925"
/translation="MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQBFLLWRMORP
GNTDAPSYSLSGTPTFTTAHMPNCHYSATLCMHANTHYWTGKMINPSCPGGLGV
TVCTWYPTQTMGSDGGVQDOAREKHVKEVISOLTRVHGTSYSPVKGDLDSLKHETLRT
HTRLVSLFNTLGLHBVSAQNPNCHICLPLNRPVSIPIVPEQWNFEINTISV
LVGLVSNLEITHSNUTCVKPSNTYTTNSQCIRWVTPPTQICVLPSPGIFVFGTSA
YRLNGSSMCFSLFIPMTIYTEODLYNYVISKPNRVRVFLIPVIGAGVGLG
TGIGITTSQFYFKLSQELNGDGMVADSLVTLQDOLNSLAAVILQNRRLDLLTAE
RGGTCFLPGECCCVYVNOGSIIVTEKVEIRDRIOREAEELRNTGPMGLLSQWMPWILP
FLAPLAIILLLGLPCIFNLVNFVSSRIEAVKLQMEPKMQSKTKIYRPLDRPASP
RSDVNDIKGPPBEISAAQPLLRNSAGSS"
misc_feature
1705..1720
/note="splice acceptor site"
LTR
1959..2694
ORIGIN
Query Match: 90.2%; Score 1335.4; DB 9; Length 2694;
Best Local Similarity 93.9%; Pred. No. 0;
Matches 1390; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
QY 1 ATGSCCTCTCCTTATCATATCTTTCTCTTTACTGCTCTTACCCCTTTGGCTCTCCT 60
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 GCACCCCTCCATGCTGCTGTACCAACCAAGTAGTCTCCCTTACCAAGAGTTTCTATGA 120
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 344 GCACCCCTCCATGCTGATGACCAAGTAGTCTCCCTTACCAAGAGTTTCTATGA 403
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 ACGGGCTTCTGGAATATTTGATGCCCATCATATATAGGAGTTTATCTAAGGAACTCC 180
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 404 ATGACGGCTCCGGAAATATTTGATGCCCATCATATAGGAGTTTCTAAGGAACTCC 463
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 181 ACCTTCACTGCCACCCATATGCCCGCACTGCTTAACTCTGCACTCTTTGATG 240
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 464 ACCTTCACTGCCACCCATATGCCCGCACTGCTTAACTCTGCACTCTTTGATG 523
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 241 CATGCAATATCTCATTTATGGACAGGAAATGATTAATCTAGTTGCTCTGGAGACTT 300
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 524 CATGCAATATCTCATTTATGGACAGGAAATGATTAATCTAGTTGCTCTGGAGACTT 583
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 301 GGAGCCACTGCTGTTGGACTTACTTCAACCATACCATATGCTGATGCTGAGGGTGGAA 360
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 584 GGAGTCACTGCTGTTGGACTTACTTCAACCATACCATATGCTGATGCTGAGGGTGGAG 643
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 361 CAAGGTCAGCAGAGAAACAGTAAGAGCAATCTCCCACTGACCCGGGACAT 420
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 644 CAAGATCAGCAGAGAAACAGTAAGAGCAATCTCCCACTGACCCGGGATCAT 703
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 421 AGCACCCCTAGCCCTACAAAGGACTAGTTCTCTCAAACTACATGAACCTCCGTAAC 480
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 704 GGCACCTCTAGCCCTACAAAGACTAGATCTCTCAAACTACATGAACCTCCGTAAC 763
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 481 CATACTCGCTGGTGGACTTATTTAATCAACCTCTACTCGGCTCCATGAGGCTCAGCC 540
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 764 CATACTCGCTGGTGGACTTATTTAATCAACCTCTACTCGGCTCCATGAGGCTCGGCC 823
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 541 CAACACCTACTACTGTTGGATGCTCCCTCGCACTTCAGGCCATACATTCATC 600
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 824 CAACACCTACTACTGTTGGATGCTCCCTCGCACTTCAGGCCATACATTCATC 883
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 601 CCGTCTCTGAACCAATGGAACCACTTCAGCAGAGAAATAAACACCACTTCGTTTTAGTA 660
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 884 CCGTCTCTGAACCAATGGAACCACTTCAGCAGAGAAATAAACACCACTTCGTTTTAGTA 943
```

```
QY 661 GGACCTCTTCTTCCATCTGGAATAAACCCATACCTCAACCTCAGCTGTGTAAATTT 720
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 944 GGACCTCTTCTTCCATCTGGAATAAACCCATACCTCAACCTCAGCTGTGTAAATTT 1003
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 721 AGCAATACTATAGACACAAACAGCTCCCAATGCATCAGGTGGGTAAACCTCCACACGA 780
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1004 AGCAATACTATAGACACAAACAGCTCCCAATGCATCAGGTGGGTAACTCTCCACACAA 1063
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 781 ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGGTACCTCAGCTATCATGTGTTG 840
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1064 ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGGTACCTCAGCTATCATGTGTTG 1123
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 841 AATGGCTCTTCAAGATCTATGTGCTTCTCTCATTTCTAGTGGCCCTATGACCATCTAC 900
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1124 AATGGCTCTTCAAGATCTATGTGCTTCTCTCATTTCTAGTGGCCCTATGACCATCTAC 1183
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 901 ACTGAACAAGATTTATACAATCATGTCTACCTAAGCCCAACAAAGAGTAGTACCATT 960
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1184 ACTGAACAAGATTTATACAATCATGTCTACCTAAGCCCAACAAAGAGTAGTACCATT 1243
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 961 CTTCTCTTCTTATCATGAGCAGGAGTGTAGGAGAGTAGTACTGCGATTTGGCAGTATC 1020
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1244 CTTCTCTTCTTATAGGAGCAGGAGTGTAGTACTGCGATTTGGCAGTATC 1303
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1021 ACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAAATAATGTGACATGGAACAG 1080
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1304 ACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAAATAATGTGACATGGAACAG 1363
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1081 GTCACTGACTCCCTGGTCACTTGTCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCTT 1140
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1364 GTGCGGACTCCCTGGTCACTTGTCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCTT 1423
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1141 CAAAATCGAAGACTTTTAGACTTTGCTAAACCCCAAGAGAGGGGAACTGTATTTTAA 1200
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1424 CAAAATCGAAGACTTTTAGACTTTGCTAAACCCCAAGAGAGGGGAACTGTATTTTAA 1483
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1201 GGAGAAGAACGCTGTTTATGTTATTAATCAATCCGAATTCGTCTAGAGAAAGTTAAAGAA 1260
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1484 GGGAGAAGAACGCTGTTTATGTTATTAATCAATCCGAATTCGTCTAGAGAAAGTTAAAGAA 1543
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1261 ATTTCGAGATCGAATACAACTAGAGAGAGAGCTTCAAAACACCGGAGCTGGGGCTC 1320
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1544 ATTTCGAGATCGAATACAACTAGAGAGAGAGCTTCAAAACACCGGAGCTGGGGCTC 1603
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1321 CTAGCCCAATGAGTGGCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1604 CTAGCCCAATGAGTGGCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1663
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1381 TTACTCTCTTTGGACCTGTATCTTTAACTCTGTTTAAAGTTTGTCTCTCCAGAAAT 1440
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1664 CTACTCTCTTTGGACCTGTATCTTTAACTCTGTTTAAAGTTTGTCTCTCCAGAAAT 1723
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1441 GAAGCTGTAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1724 GAAGCTGTAAGCTACAAATGGAACCCCAAGATGCAATCCAA 1764
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 20
AF520487S2
LOCUS 2694 bp DNA linear PRI 11-FEB-2004
DEFINITION Homo sapiens individual 148 allele B, envelope glycoprotein gene,
complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION AF520488
VERSION AF520488.1 GI:33410944
KEYWORDS 2 of 2
SEGMENT Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2694)
AUTHORS Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
```

TITLE The endogenous retroviral locus ERWE1 is a bona fide gene involved in hominoid placental physiology
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
PUBMED 14757826
REFERENCE 2 (bases 1 to 2694)
AUTHORS Mallet, F., Bouton, O. and Oriol, G.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2002) Retrovirology Department, UMR 2142 CNRS-bioMérieux, Ecole Normale Supérieure de Lyon - 46 allée d'Italie, Lyon 69364 cedex 07, France
FEATURES
source
1..2694
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="individual 148 allele B"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-q22"
/sex="male"
/cell_type="BMC's"
/note="Caucasian
endogenous virus: HERV-W"
misc_feature
43..57
/note="splice acceptor site"
284..1900
/note="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AA017566.1"
/db_xref="GI:33410945"
/translation="MALPYHIFLFTLLPSFTLTAPPKRCMTSSSPYQEFILWRMQR
GNIDAPSVLSKGTPTFTAHTRPNCHYSAITLCHMANTHYWTKMINPSCPGLGV
TVCWTFPTQMGSDGGVQDAEKVKEVISOLTRVHGTSVPVKGLDLSKLHETLRT
HTRLSVFNLTLLHEVSAQNPTNCWICLPNRPVYSIPVPEQNNSTEINTSV
LVGLVSNLTITSLNLTCKFSNTYTNNSQCIRWTPPTQIVLPSGIFVCGISA
YRLNGSSMCSFLSPVPMYIYEQDLNYYVISKPRKRVPLFPFVIGALG
TGIGITTSFOFYKLSQELNGDMRVADSLVLOQLNSLAALVQNRRLDLLTAE
RGCTPLFGRECCYVYNGSIVTEKVEIKRDIQRRAELRNTGPNGLLSOMWPLIP
FLGLAAILLLLEFGPCIFNLVNFVSSRIEAVKLQWEPKQSKTKIYRPLDRPASP
RSDVNDIKGIPPEISIAQPLLRNSAGSS"
misc_feature
1705..1720
/note="splice acceptor site"
1959..2694
LTR
ORIGIN
Query Match 90.2%; Score 1335.4; DB 9; Length 2694;
Best Local Similarity 93.9%; Pred. No. 0;
Matches 1390; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
QY 1 ATGGCCCTCCCTATCATCTTTCTCTTTACTGTTCTCTTACCCCTTTGGCTCTCACT 60
DB 284 ATGGCCCTCCCTATCATATTTTCTCTTTACTGTTCTTTTACCCTCTTTCACCTCACT 343
QY 61 GCACCCCTCCATGCTGTGTACACACAGTAGCTCCCTTACCAAGAGTTTCTATGAAGA 120
DB 344 GCACCCCTCCATGCTGTGTATGACAGTAGTCCCTTACCAAGAGTTTCTATGAGA 403
QY 121 ACGGGCTTCTGGNAATATTGATGCCCATCATATAGGAGTTTATCTAAGGGAACTCC 180
DB 404 ATGCAGCGTCCGGAAATATTGATGCCCATCATATAGGAGTTTCTAAGGGAACTCC 463
QY 181 ACCTTCACTGCCACACCATATGCCCGCAACTGCTATAACTCTGCGCACTCTTTGCAATG 240
DB 464 ACCTTCACTGCCACACCATATGCCCGCAACTGCTATCACTCTGCGCACTCTTTGCAATG 523
QY 241 CATGCAAACTCATATTATGACAGGGAATGATTAATCTAGTTGCTCTGGAGGACTT 300
DB 524 CATGCAAACTCATATTATGACAGGGAATGATTAATCTAGTTGCTCTGGAGGACTT 583
QY 301 GGAGCCACTGCTGTGGACTTACTTACCACATACAGTAGTCTGTATGCTGGGGTGGAAAT 360
DB 584 GGAGTCACTGCTGTGGACTTACTTACCACATACAGTAGTCTGTATGCTGGGGTGGAGTT 643

QY 361 CAAGCTCAGGCAAGAGAAAAACAAGTAAAGAGCAATCTCCCACTGACCCGGGGCAT 420
DB 644 CAAGTACAGGCAAGAGAAAAACAATGTAAGAGAAATCTCCCACTGACCCGGGGTAT 703
QY 421 AGCAGCCCTAGCCCTACAAAGGACTAGTCTCTCAAACTACATGAAACCCCTCGTACC 480
DB 704 GGCACCTCTAGCCCTACAAAGGACTAGTCTCTCAAACTACATGAAACCCCTCGTACC 763
QY 481 CATATCTCGCTGCTGAGCCTATTATTAATACCACTCTCTCGGCTCCATGAGTCTCAGC 540
DB 764 CATATCTCGCTGCTGAGCCTATTATTAATACCACTCTCTCGGCTCCATGAGTCTCGGCC 823
QY 541 CAAAACCTCTACTAACTGTTGGATGTCCTCCCTGCACTTCAGGCCCATATTTCAATC 600
DB 824 CAAAACCTCTACTAACTGTTGGATGTCCTCCCTGCACTTCAGGCCCATATTTCAATC 883
QY 601 CTTGTTCTCGAACAATGGAACAATCTCAGCAGAGAAATAACACCACTTCCCTTTTAGTA 660
DB 884 CTTGTTCTCGAACAATGGAACAATCTCAGCAGAGAAATAACACCACTTCCCTTTTAGTA 943
QY 661 GGACCTCTTTGTTTCCAATCTGGAATAACCCATACCTCAAACTCACCTGTGTAAAAATTT 720
DB 944 GGCCTCTTTGTTTCCAATCTGGAATAACCCATACCTCAAACTCACCTGTGTAAAAATTT 1003
QY 721 AGCAATCTATAGACACAAACAGCTCCCAATGCATCAGTGGGTAACTCCCAACGA 780
DB 1004 AGCAATCTATACATACAAACCACTCCCAATGCATCAGTGGGTAACTCCCTCCACAA 1063
QY 781 ATAGTCTGCTACCTCCCTCAGGAATATTTTGTCTGTGTGCTACCTCAGCTCATCATTTGTTG 840
DB 1064 ATAGTCTGCTACCTCCCTCAGGAATATTTTGTCTGTGTGCTACCTCAGCTCATCATTTGTTG 1123
QY 841 AATGGCTCTTCAGAACTATATGTCCTCTCATTTAGTGCCTCTATGACCATCTAC 900
DB 1124 AATGGCTCTTCAGAACTATATGTCCTCTCATTTAGTGCCTCTATGACCATCTAC 1183
QY 901 ACTGAAACAAGATTATACATCATGTCTACCTAGCCCCCAACAAAAGAGTACCATT 960
DB 1184 ACTGAAACAAGATTATACATCATGTCTATCTAAGCCCCCAACAAAAGAGTACCATT 1243
QY 961 CTTCTCTTTGTTATCAGACAGAGTGTCTAGGCAGACTAGTCTAGGCACTGGCACTATC 1020
DB 1244 CTTCTCTTTGTTATAGACAGAGTGTCTAGGCACTAGGCACTGGCACTATC 1303
QY 1021 ACAACCTCTACTCAGTCTTACTACAAATCTCTCAAGAAATAAATGTCACATGGAACAG 1080
DB 1304 ACAACCTCTACTCAGTCTTACTACAAATCTCTCAAGAAATAAATGTCACATGGAACAG 1363
QY 1081 GTCACTGACTCCCTGCTCAGCTTGCAGATCAACTTAACTCCCTAGCAGTAGTCCCTT 1140
DB 1364 GTCCGCACTCCCTGCTCAGCTTGCAGATCAACTTAACTCCCTAGCAGTAGTCCCTT 1423
QY 1141 CAAAATCGAAGAGCTTTAGACTTTGTAAACCGCAAGAGGGGAAACCTGTTTATTTTA 1200
DB 1424 CAAAATCGAAGAGCTTTAGACTTTGTAAACCGCTGAAAGAGGGGAAACCTGTTTATTTTA 1483
QY 1201 GGAAGAAAGCTGTTTATTTATTTATCAATCCAGATTTGCTCACTGAGAAAGTTAAAGAA 1260
DB 1484 GGAAGAAAGCTGTTTATTTATTTATTTATCAATCCAGATTTGCTCACTGAGAAAGTTAAAGAA 1543
QY 1261 ATTTCAGATCGAATACAAATGTAGACAGAGAGCTTCAAAACACCAACCTCGGGGCTC 1320
DB 1544 ATTTCAGATCGAATACAAATGTAGACAGAGAGCTTCAAAACACCAACCTCGGGGCTC 1603
QY 1321 CTGAGCAATAGGATGCTCCCTGCTTCTCCCTTCTTAGGACTCTAGAGCTCTAATATTG 1380
DB 1604 CTGAGCAATAGGATGCTCCCTGCTTCTCCCTTCTTAGGACTCTAGAGCTCTAATATTG 1663
QY 1381 TTACTCTCTTTGGAACCTGTATCTTTAACTCTCTGTTTAACTTTGCTCTTCCAGAAAT 1440
DB 1664 CTACTCTCTTTGGAACCTGTATCTTTAACTCTCTGTTTAACTTTGCTCTTCCAGAAAT 1723
QY 1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATAGGAACCCCA 1481

Db	1724	GAAGCTGTAAACTACAAATGGAGCCCAAGATGCAATGCCAA	1764
RESULT 21	AF520489S2	2694 bp DNA linear	PRI 11-FEB-2004
LOCUS	Homo sapiens individual 22 allele A, envelope glycoprotein gene,		
DEFINITION	complete cdb, and 3' long terminal repeat, complete sequence.		
ACCESSION	AF520490		
VERSION	AF520490.1	GI:33410948	
KEYWORDS			
SEGMENT	2 of 2		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	1 (bases 1 to 2694)		
TITLE	Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G.,		
JOURNAL	Bonnaud, B., Lucotte, G., Duret, L., and Mandrand, B.		
PUBLISHED	The endogenous retroviral locus ERVWE1 is a bona fide gene involved		
REFERENCE	in hominoid placental physiology		
AUTHORS	Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)		
JOURNAL	14757826		
TITLE	2 (bases 1 to 2694)		
AUTHORS	Mallet, F., Bouton, O., and Oriol, G.		
JOURNAL	Direct Submission		
FEATURES	Submitted (07-JUN-2002) Retrovirology Department, UMR 2142		
source	CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée		
	d'Italie, Lyon 69364 cedex 07, France		
	Location/Qualifiers		
	1..2694		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/isolate="individual 22 allele A"		
	/db_xref="taxon:9606"		
	/chromosome="7"		
	/map="7q21-q22"		
	/sex="male"		
	/cell_type="PBMC's"		
	/notes="Caucasian"		
	endogenous_virus: HERV-W"		
misc_feature	43..57	/notes="splice acceptor site"	
CDS	284..1900	/notes="syncytin"	
	/codon_start=1		
	/product="envelope glycoprotein"		
	/protein_id="AAQ17567.1"		
	/db_xref="GI:33410949"		
	/translation="MALPYHIFLFTVLPLSFTLAPPCCRMSTSSPYQEFLLWRMORP		
	GNIDAPSYRSLSGTPTTAHTMPRNCYHSATLCMHANTHYMTGRMNPSPCGGLGV		
	TVCTWYFTTGMSDGGVDOAREKHVKEVISOLTRVHGTSPPYKGLDLSKHLTLRT		
	HYPLVSLFTLTLGLHVSANQNTNCWICLPLNFRPYVSIPIVPEQWNSTEINTSV		
	LYGLVSNLEHTSLNLTCKVSNITTYTNSQCIRWTPPTQVLCPLSGIFVFCVGTSA		
	XVCLNGSSSMETFLVLPMTIYTDLYNIVISYRKNRPVILPFIYVAGVYLGALG		
	TGGTGTSTQFYKLSQELNGEMERVADSLVTLQDLNSLAALVQLNRRALDLTAIE		
	RGCTCLFLGEBCCYVYVQSGIVTEKVEIRDRIQRRAEIRNTGPMGLLSQWMPWTLF		
	FLGPLAAIILLGLFGPCIFNLNVFVSSRIEAVKLQWPKMQSKTKIYRPLDRPAS		
	RSVDNDIKGTPEEISAAQPLLRPSAGSS"		
misc_feature	1705..1720		
LTR	1959..2694	/notes="splice acceptor site"	
ORIGIN			
Query Match	90.2%	Score 1335.4; DB 9; Length 2694;	
Best Local Similarity	93.94;	Pred. No. 0;	
Matches 1390; Conservative	0; Mismatches	91; Indels	0; Gaps
QY	1	ATGCCCTCCCTTATCATATTTTCTTTTACGTGTTCTCTTACCCCTTTCCGCTCTCACT	60
Db	284	ATGCCCTCCCTTATCATATTTTCTTTTACGTGTTCTTTTACCTCTTCTCACTCTCACT	343

QY	61	GCACCCCTCCATGCTGCTGTACAAACAGTAGCTCCCTTACCAAGAGTTTCTATGAAGA	120
Db	344	GCACCCCTCCATGCTGCTGTAGCAAGTAGCTCCCTTACCAAGAGTTTCTATGAGA	403
QY	121	ACGCGGCTTCCTGGAATATTGATGCCCATCATATAGGAGTTTATTAAGGGGAACTCC	180
Db	404	ATGAGGGTCCCGGAATATTGATGCCCATCGTATAGGAGTTTCTTAAGGGNACCCCC	463
QY	181	ACCTTCACTGCCCCACACCCCATATGCCCGCAACTGCTATTAACCTCTGTCATCTTTG	240
Db	464	ACCTTCACTGCCCCACACCCCATATGCCCGCAACTGCTATTAACCTCTGTCATCTTTG	523
QY	241	CATGCAATACTCATTTATTTGGACAGGGAATGATTAATCTAGTTGCTCGGAGGACTT	300
Db	524	CATGCAATACTCATTTATTTGGACAGGGAATGATTAATCTAGTTGCTCGGAGGACTT	583
QY	301	GGAGCCACTGCTGTTGGACTTACTTCAACCATACAGATATGCTGATGGGGTGGAAAT	360
Db	584	GGAGTCACTGCTGTTGGACTTACTTCAACCATGCTGATGCTGATGGGGTGGAGTT	643
QY	361	CAAGGTGAGCAAGAGAAAAAAGTAAGGAAGCAATCTCCCAACTGACCCCGGGACAT	420
Db	644	CAAGATCAGCAAGAGAAAAAAGTAAGGAAGCAATCTCCCAACTGACCCCGGGACAT	703
QY	421	AGCACCCCTAGCCCTACAAAGGACTAGTTCTCTCAAACTACATGAACCCCTCCGTACC	480
Db	704	GGCACCTCTAGCCCTACAAAGGACTAGATCTCTCAAACTACATGAACCCCTCCGTACC	763
QY	481	CATACCTCGCTGCTGAGCCCTATTTAATACCAACCTCACTCGGCTCCATGAGGTCACGC	540
Db	764	CATACCTCGCTGCTGAGCCCTATTTAATACCAACCTCACTCGGCTCCATGAGGTCACGC	823
QY	541	CAAAACCTACTAACTGTTGGATGTGCTCCCTCCCTGACCTTTCAGGCCATACATTTCAATC	600
Db	824	CAAAACCTACTAACTGTTGGATGTGCTCCCTCCCTGACCTTTCAGGCCATACATTTCAATC	883
QY	601	CTGTTCCTGAAACAATGGAAACAATTCAGCAGAGAAATAAACCACCTCTCCGTTTAGTA	660
Db	884	CTGTACCTGAAACAATGGAAACAATTCAGCAGAGAAATAAACCACCTCTCCGTTTAGTA	943
QY	661	GGACCTCTTGTTCCTCAATCTGGAATAAACCCATACCTCAAACTCTGTTGTTAAATTT	720
Db	944	GGACCTCTTGTTCCTCAATCTGGAATAAACCCATACCTCAAACTCTGTTGTTAAATTT	1003
QY	721	AGCAATACTATAGACAACAACCACTCCCAATGATCAGGTGGGTAAACCTCCACACGA	780
Db	1004	AGCAATACTATAGACAACAACCACTCCCAATGATCAGGTGGGTAAACCTCCACACGA	1063
QY	781	ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGTTACCTCAGCCTATCATTTGTTTG	840
Db	1064	ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGTTACCTCAGCCTATCATTTGTTTG	1123
QY	841	AATGGCTCTTCAAGATCTATGCTTCTCTCAATCTTCTAGTCCCTTATGACCATCTAC	900
Db	1124	AATGGCTCTTCAAGATCTATGCTTCTCTCAATCTTCTAGTCCCTTATGACCATCTAC	1183
QY	901	ACTGAAACAAGATTTATACAATCATGCTGCTACCTTAAGCCCCACAAACAAAGAGTACC	960
Db	1184	ACTGAAACAAGATTTATACAATCATGCTGCTACCTTAAGCCCCACAAACAAAGAGTACC	1243
QY	961	CTTCTCTTTTGTATACAGAGAGGTGCTAGGAGACTAGGTACTGGCATTTGGAGTATC	1020
Db	1244	CTTCTCTTTTGTATACAGAGAGGTGCTAGGAGACTAGGTACTGGCATTTGGAGTATC	1303
QY	1021	AGAACCTCTACTCAGTTCTTACTACAACTATCTCAAGAAATAAATCGTGACATGGACAG	1080
Db	1304	AGAACCTCTACTCAGTTCTTACTACAACTATCTCAAGAAATAAATCGTGACATGGACAG	1363
QY	1081	GTCACTGACTCCCTGCTGCTTGGCAAGATCAACTTAACTCCCTAGCAGAGTAGTCCCT	1140
Db	1364	GTCCGCGACTCCCTGCTGCTTGGCAAGATCAACTTAACTCCCTAGCAGAGTAGTCCCT	1423
QY	1141	CAAAATCGAAGAGCTTTAGACTTGTCTAACGCCCAAAAGAGGGGAAACCTGTTTATTTTA	1200

Db	1424	CAAAATCGAAGAGCTTTAGACCTTCTAACCGCTCAAGAGGGGGAACCTGTTTATTTT	1483	
Qy	1201	GGAGAAGAACGCTGTTTATTTATGTTAATCAATCCAGAAATCTCTACGAGAAAGTTAAAGAA	1260	
Db	1484	GGGGAAGAATCTGTTTATTTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA	1543	
Qy	1261	ATTGAGATCGAATACAAATGTAGACAGAGAGCTTCAAAACACCGAACGCTGGGCGCTC	1320	
Db	1544	ATTGAGATCGAATACAAATGTAGACAGAGAGCTTCAAAACACCGAACGCTGGGCGCTC	1603	
Qy	1321	CTGACCAATGAGTGCCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG	1380	
Db	1604	CTGACCAATGAGTGCCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG	1663	
Qy	1381	TTACTCCTCTTTGGACCTGTATCTTTAACTCCTCTGTTAAAGTTGTCTCTCCAGAAAT	1440	
Db	1664	CTACTCCTCTTTGGACCTGTATCTTTAACTCCTCTGTTAAAGTTGTCTCTCCAGAAAT	1723	
Qy	1441	GAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA	1481	
Db	1724	GAGCTGTAAAGCTACAAATGAGCCCAAGATGAGTCCCA	1764	
RESULT 22				
AF520505S2				
LOCUS				
DEFINITION	AF520505S2 2694 bp DNA linear PRI 11-FEB-2004			
ACCESSION	Homo sapiens individual 37 allele A, envelope glycoprotein gene,			
VERSION	complete cds, and 3' long terminal repeat, complete sequence.			
KEYWORDS	AF520506			
SEGMENT	AF520506.1 GI:33410980			
SOURCE	2 of 2			
ORGANISM	Homo sapiens (human)			
REFERENCE	Homo sapiens			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
JOURNAL	1 (bases 1 to 2694)			
PUBMED	Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G.,			
AUTHORS	Bonnaud, B., Lucotte, G., Duret, L. and Mandrand, B.			
TITLE	The endogenous retroviral locus ERVWE1 is a bona fide gene involved			
JOURNAL	in hominoid placental physiology			
PUBMED	Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)			
REFERENCE	14757826			
AUTHORS	2 (bases 1 to 2694)			
TITLE	Mallet, F., Bouton, O. and Oriol, G.			
JOURNAL	Direct Submission			
PUBMED	Submitted (07-JUN-2002) Retrovirology Department, UMR 2142			
FEATURES	CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée			
source	d'Italie, Lyon 69364 cedex 07, France			
	Location/Qualifiers			
	1..2694			
	/organism="Homo sapiens"			
	/mol_type="genomic DNA"			
	/isolate="individual 37 allele A"			
	/db_xref="taxon:9606"			
	/chromosome="7"			
	/map="7q21-q22"			
	/sex="male"			
	/cell_type="PBMC's"			
	/note="endogenous_virus: HERV-W"			
misc_feature	43..57			
	/note="splice acceptor site"			
CDS	284..1900			
	/note="syncytin"			
	/codon_start=1			
	/product="envelope glycoprotein"			
	/protein_id="AAQ17575.1"			
	/db_xref="GI:33410981"			
	/translation="MALPVHFLFTVLLPSFTLTAPPRCRWTSSTSSPYQEFILWRMQR			
	GNIDAPSRSLSKGTFPTAHTMPKNCVHSATLCMHANTHWYTKMKNPSPGGLGV			
	TVCTWYFTQGMDSGGVQDQAREKHEVLSOLTRVHGTSIPYKGLDLSKLHLRT			
	TCTHEVNSAQNPTNCWICLPLNFRPYSIPVPEQWNNFSTEINTSV			

```
Db 1124 AATGGCTCTTCAGAACTCTATGTGCTTCTCTCATCTTCTAGTGCCCTATGACCATCTAC 1183
Qy 901 ACTGAACAAGATTATATCAATCATGTGCTACTTAAGCCCAACAAAGAGTAGTACCATT 960
Db 1184 ACTGAACAAGATTATATCAATCATGTGCTACTTAAGCCCAACAAAGAGTAGTACCATT 1243
Qy 961 CTTCCCTTTCTTATCAGACAGAGTGCTAGGAGACTAGTACTGGCATTTGGCAGTATC 1020
Db 1244 CTTCCCTTTCTTATGAGACAGAGTGCTAGGAGACTAGTACTGGCATTTGGCAGTATC 1303
Qy 1021 ACAACCTCTACTCAGTTCTACTACAACTACTCTCAAGAAATAAATGGTGTGACATGGAACAG 1080
Db 1304 ACAACCTCTACTCAGTTCTACTACAACTACTCTCAAGAAATAAATGGGACATGGAACG 1363
Qy 1081 GTCACTGACTCCCTGGTCACTTGGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCCT 1140
Db 1364 GTCCCGACTCCCTGGTCACTTGGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCCT 1423
Qy 1141 CAAAATCGAAGAGCTTTAGACTTGTCTAACCCTCAAGAGAGGGGAACTCTTTATTTTA 1200
Db 1424 CAAAATCGAAGAGCTTTAGACTTGTCTAACCCTCAAGAGAGGGGAACTCTTTATTTTA 1483
Qy 1201 GGAGAAGAACGCTGTTATTTAATCAATCCAGAAATGCTCACTGAGAAAGTTAAAGAA 1260
Db 1484 GGGGAAGAACTGTTATTTAATCAATCCGAAATGCTCACTGAGAAAGTTAAAGAA 1543
Qy 1261 ATTCGAGATCGAATACAACTAGACAGAGAGCTTCAAAACACCGAACCTGGGGCCTC 1320
Db 1544 ATTCGAGATCGAATACAACTAGACAGAGAGCTTCAAAACACCTGGAGCCTGGGGCCTC 1603
Qy 1321 CTGAGCCAAATGGATGCCCTGGTCTCCCTCTTAGGACCTTAGCAGCTCTAATATTG 1380
Db 1604 CTGAGCCAAATGGATGCCCTGGTCTCCCTCTTAGGACCTTAGCAGCTCTAATATTG 1663
Qy 1381 TTAATCTCTTTGACCCCTGTATCTTTAACTCTCTTAACTTTGCTCTCCAGAAAT 1440
Db 1664 CTACTCTCTTTGACCCCTGTATCTTTAACTCTCTTAACTTTGCTCTCCAGAAATC 1723
Qy 1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
Db 1724 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1764

RESULT 23
AF520517S2
LOCUS
DEFINITION
Homo sapiens individual 55 allele A, envelope glycoprotein gene,
complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION
AF520518
VERSION
AF520518.1
KEYWORDS
GI:33411004
SEGMENT
2 of 2
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2694)
Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
The endogenous retroviral locus ERVW1 is a bona fide gene involved
in hominoid placental physiology
Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
JOURNAL
14757826
PUBMED
2 (bases 1 to 2694)
REFERENCE
Mallet,F., Bouton,O. and Oriol,G.
Direct Submission
AUTHORS
Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
CNRS-bioMérieux, Ecole Normale Supérieure de Lyon - 46 allée
d'Italie, Lyon 69364 cedex 07, France
JOURNAL
Location/Qualifiers
FEATURES
source
1..2694
/organism="Homo sapiens"
/mol_type="genomic DNA"
```

```
misc_feature
43..57
/notes="splice acceptor site"
CDS
284..1900
/notes="syncytin"
/notes="envelope glycoprotein"
/protein_id="AAQ17581.1"
/db_xref="GI:33411005"
/translation="MALPYHIFLFTLVLPSTLAPPKRCMTSSSPYQBFLLWRMQR
GNIDAPSYRSLSKGTPTFTATHMPNRCVHSATLCMHANTHYWTGMINSPCGGLGV
TWCTVFTTGMSDGGVDOAREKHVKEVISOLTRVHGTSSPYKGLDLSKLHETLRT
HTRLSLFNTLTGLHEVSAQNPNTNCMI CLPLNFRPVSI PVPKWNNESTENTTSV
LVGPLVSNLEI THSNLTCCVKSNTTYTNSQCIRWVPTPTQIVCLPSGFFVCGTSA
YRLNGSSSMCLFSLFVPPMTIYTDQDLSNVISKPNKRVPILFPVIGVILGALG
FGTCLFLGECYVYVQSGIVTEKVEIRDRIORAEELRNTGPKGLLSOMMPWLTAE
RGCLAAIILLLLFGPCIENLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRLPDRPASP
RSDVNDIKCTPPEEISAAQPLLRPNAGSS"
misc_feature
1705..1720
/notes="splice acceptor site"
LTR
1959..2694
ORIGIN

Query Match 90.2%; Score 1335.4; DB 9; Length 2694;
Best Local Similarity 93.9%; Pred. No. 0;
Matches 1390; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

Qy 1 ATGGCCCTCCCTTATCATCTTTCTTTCTTACTTGTTCTTACCCCTTTCCGCTCTCACT 60
Db 284 ATGGCCCTCCCTTATCATATTTTCTTTACTTGTTCTTTTACCCCTTTCTCACTCTCACT 343

Qy 61 GCACCCCTCCATGCTGTGACCAACAGTAGTCTCCCTTACCAAGAGTTTCTATGAAGA 120
Db 344 GCACCCCTCCATGCTGTATGACAGTAGTCTCCCTTACCAAGAGTTTCTATGAAGA 403

Qy 121 ACGCGCTTCTGGAAATATTGATGCCCATCATATAGGATTTATCTAAGGAAACTCC 180
Db 404 ATGACACGCTCCGAAATATTGATGCCCATCATATAGGAGTCTTCTTAAAGGAACCC 463

Qy 181 ACCTTCACCTGCCACACCCATATGCCCCGCACTGCTATAACTCTGCACTCTTTGCATG 240
Db 464 ACCTTCACCTGCCACACCCATATGCCCCGCACTGCTATACTCTGCACTCTTTGCATG 523

Qy 241 CATGCAAACTACTCATTTATTGGACAGGAAATGAATTAATCTTAGTTGTCTCGGAGACTT 300
Db 524 CATGCAAACTACTCATTTATTGGACAGGAAATGAATTAATCTTAGTTGTCTCGGAGACTT 583

Qy 301 GGAGCAGCTGTCTGTGGACTTACTTACCCATACCACTAGTATGCTCATGGGGTGGNATT 360
Db 584 GGAGTCACTGTCTGTGGACTTACTTACCCAACTGGTATGCTCATGGGGTGGAGTT 643

Qy 361 CAAGTGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Db 644 CAAGATCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 703

Qy 421 AGCACCCCTAGCCCTTACAAAGAGCTAGTTCTCTCAAAACTACATGAAACCTCCCGTACC 480
Db 704 GGCACCTCTAGCCCTTACAAAGAGCTAGATCTCTCAAAACTACATGAAACCTCCCGTACC 763

Qy 481 CATACTCGGCTGTGAGCCCTATTATATACACCCCTCACTCGGCTCCATGAGGTCTACGCC 540
Db 764 CATACTCGGCTGTGAGCCCTATTATATACACCCCTCACTCGGCTCCATGAGGTCTCGGCC 823

Qy 541 CAAAACCTACTAACTGTTGGATGTGCTCCCTCCCTGCATCTTCAAGGCCATACATTTCATC 600
Db 824 CAAAACCTACTAACTGTTGGATGTGCTCCCTCCCTGCATCTTCAAGGCCATATGTTTCAATC 883
```


301 GGAGCACTGCTGTTGGACTTACTTCCACCATACCACTATGCTGATGGGGTGAATT 360
584 GGAGTCACTGCTGTTGGACTTACTTCCACCAACTGATGCTGATGGGGTGAATT 643
361 CAAGGTCAGGCAAGAGAAAAACAAGTAAGAGAACTATCTCCCAACTGACCCGGGACAT 420
644 CAAGATCAGGCAAGAGAAAAACAATGTAAGAAAGTAATCTCCCAACTCACCCGGGTACAT 703
421 AGACCCCTAGCCCTTCAAGAGACTAGTTCTCTCAAACTACATGAACCCCTCGTACC 480
704 GGCACCTCTAGCCCTTCAAGAGACTAGATCTCTCAAACTACATGAACCCCTCGTACC 763
481 CATACTCGCTGCTGAGCCCTATTATATACCACTCTGAGTCTGAGTCTGAGTCTGAGCC 540
764 CATACTCGCTGCTGAGCCCTATTATATACCACTCTGAGTCTGAGTCTGAGTCTGAGCC 823
541 CAAAACCTACTAATCTGTTGGATGCTGCTCCCTGCTGACCTTTCAGGCCATACATTTCAATC 600
824 CAAAACCTACTAATCTGTTGGATGCTGCTCCCTGCTGACCTTTCAGGCCATATGTTCAATC 883
601 CTTGTTCTTGAACATGGAACAACTTTCAGCACAGAAATAAACCACTTCCGTTTGTAGTA 660
884 CTTGTTCTTGAACATGGAACAACTTTCAGCACAGAAATAAACCACTTCCGTTTGTAGTA 943
661 GGACCTCTGTTTCCCAATCTGGAATTAACCACTCTCAACCTCTGAGTCTGTAATTTT 720
944 GGACCTCTGTTTCCCAATCTGGAATTAACCACTCTCAACCTCTGAGTCTGTAATTTT 1003
721 AGCAATACTATAGACAAACAGCTCCCAATGATCAGGTGGGTAAACCTCCCAACAGA 780
1004 AGCAATACTATAGACAAACAGCTCCCAATGATCAGGTGGGTAAACCTCCCAACAGA 1063
781 ATAGTCTGCTTACCTCAGGAATATTTTGTCTGTTGTTACCTCAGCTATCATATGTTG 840
1064 ATAGTCTGCTTACCTCAGGAATATTTTGTCTGTTGTTACCTCAGCTATCATATGTTG 1123
841 AATGGCTCTTCAGAACTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAC 900
1124 AATGGCTCTTCAGAACTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAC 1183
901 ACTGAACAGATTTATACAACTATGCTGCTTACCTTAAAGCCCAACAAAGAGTACCCATT 960
1184 ACTGAACAGATTTATACAACTATGCTGCTTACCTTAAAGCCCAACAAAGAGTACCCATT 1243
961 CTTCTCTTTTGTATCAGAGCAGGAGTCTAGGAGACTAGTACTGCGCATTTGCGAGTATC 1020
1244 CTTCTCTTTTGTATCAGAGCAGGAGTCTAGGAGACTAGTACTGCGCATTTGCGGATATC 1303
1021 ACAACCTCTACTCAGTCTTACTACAACTATCTCAAGAAATAAATGCTGATGGAACAG 1080
1304 ACAACCTCTACTCAGTCTTACTACAACTATCTCAAGAAATAAATGCTGATGGAACAG 1363
1081 GTCACCTGACCTCCCTGCTGCTTGCAGATCAACTTAACTCCCTAGCAGCAGTAGTCCCTT 1140
1364 GTCGCGCATCTCCCTGCTGCTTGCAGATCAACTTAACTCCCTAGCAGCAGTAGTCCCTT 1423
1141 CAAAATCGAAGAGCTTTAGACTTGTCTAAGCCCAAGAGAGGGGGAACCTGTTTATTTTA 1200
1424 CAAAATCGAAGAGCTTTAGACTTGTCTAAGCCCAAGAGAGGGGGAACCTGTTTATTTTA 1483
1201 GGAGAGAACGCTGTTATATGTTATTAATCAATCAGAAATTTGTCATGAGAAAGTTAAAGAA 1260
1484 GGGGAGAAATGCTGTTATATGTTATTAATCAATCAGAAATTTGTCATGAGAAAGTTAAAGAA 1543
1261 ATTGAGATCGAATACAACTAGCAGAGAGGAGCTTCAAAACACCGAAGCTGGGCGCTC 1320
1544 ATTGAGATCGAATACAACTAGCAGAGAGGAGCTTCAAAACACCGAAGCTGGGCGCTC 1603
1321 CTGAGCAATGAGTGGCTGCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAT 1380
1604 CTGAGCAATGAGTGGCTGCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAT 1663
1381 TTACTCCTCTTTGGACCCCTGATCTTTTAACTCCTCTGTTAAAGTTTGTCTCTCCAGAAAT 1440

Db 1664 CTTACTCTCTTTGGACCTGTATCTTTAACTCTCTGTTAACTTTGTTCTCTCCAGATC 1723
Qy 1441 GAAGCTGTAAAGCTACAGATGGTCTTTACAAATGGAACCCCA 1481
Db 1724 GAAGCTGTAAAGCTACAAATGGAGCCCAAGATGAGTCCAA 1764
RESULT 25
AF520521S2 2694 bp DNA linear PRI 11-FEB-2004
LOCUS Homo sapiens individual 6 allele A, envelope glycoprotein gene,
DEFINITION complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION AF520522
VERSION AF520522.1 GI:33411012
KEYWORDS 2 of 2
SEGMENT Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2694)
AUTHORS Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
TITLE The endogenous retroviral locus ERVWE1 is a bona fide gene involved
in hominoid placental physiology
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
PUBMED 14757826
REFERENCE 2 (bases 1 to 2694)
AUTHORS Mallet,F., Bouton,O. and Oriol,G.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée
d'Italie, Lyon 69364 cedex 07, France
FEATURES
1..2694
source
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="individual 6 allele A"
/db_xref="taxon:9606"
/chromosomes="7"
/map="7q21-q22"
/sex="male"
/cell_type="PBMC's"
/notes="Caucasian"
endogenous_virus: HERV-W
misc_feature 43..57
Location/Qualifiers
/note="splice acceptor site"
CDS 284..1900
/notes="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAO17583.1"
/db_xref="GI:33411013"
/translation="MALPFIHFLFTVLLPSFLTAPPCCRCMTSSSPYQBLRMQRPP
GNIDAPYSKSLSGTFTFTAHMPNRYHSATLCMHANTHYWTGRMINSPCGGLGV
TVCWTYFTTGMGSDGQDQAREKHVEVISQLTRVHGTSPPYKGLDLSKLHRLT
HTRLVSLFTLTLLTGLHEVSAQNPTNCWICLPLNFRPVSVIPVPEQNNFSTEINTTSV
LVGLAVSLNLEITHTSNLTCTKPSNTTNTTNSQICRWTPPTQIVCLPSGIFPVCGTSA
YRLANGSSRSMCFSLPVPMTLYTEODLYNVIKPRNRPVLPVPIVAGVGLGALG
TIGGTTSTQFYKLUSBLNGDMERVADSLVTQLQSLAAVVLQNRALLLTAE
RGCTCLFLGEECCYVNSQGI VTEKVKIEIRDRIRAEELRNTGPMGLLSQWMPWILP
FLGPLAAIILLLLFGPCI FNLLVNFVSSRIEAVKLQMEPRMQSKTKIYRRLRPLRASP
RDVNDIKGTPPBEISAAQPLLRPNASGSS"
misc_feature 1705..1720
/note="splice acceptor site"
LTR 1959..2694
ORIGIN
Query Match 90.2%; Score 1335.4; DB 9; Length 2694;
Best Local Similarity 93.9%; Pred. No. 0;
Matches 1390; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

```
QY 1 ATGGCCCTCCCTTATCATATCTTTCTCTTTACTGTTCTCTTACCCCTTTGCTCTCACT 60
Db |||||
QY 284 ATGGCCCTCCCTTATCATATCTTTCTCTTTACTGTTCTCTTACCCCTTTGCTCTCACT 343
Db |||||
QY 61 GCACCCCTCCCTTATCATATCTTTCTCTTTACTGTTCTCTTACCCCTTTGCTCTCACT 120
Db |||||
QY 344 GCACCCCTCCCTTATCATATCTTTCTCTTTACTGTTCTCTTACCCCTTTGCTCTCACT 403
Db |||||
QY 121 AGCGGCTCTCTGGAATATTTGATGCGCCATCATATAGGAGTTTCTTAAGGAACTCC 180
Db |||||
QY 404 ATGACGCTCCCGGAAATATTTGATGCGCCATCATATAGGAGTTTCTTAAGGAACTCC 463
Db |||||
QY 181 ACCTTCACCTGCCACACCCATATGCGCCGCACTGCTATTAACCTCTGCCACTCTTTGCATG 240
Db |||||
QY 464 ACCTTCACCTGCCACACCCATATGCGCCGCACTGCTATTAACCTCTGCCACTCTTTGCATG 523
Db |||||
QY 241 CATGCAAACTACTCATTTATTTGGACAGGAAATGATTAATCTTAGTTGTTCTTGAGGACTTT 300
Db |||||
QY 524 CATGCAAACTACTCATTTATTTGGACAGGAAATGATTAATCTTAGTTGTTCTTGAGGACTTT 583
Db |||||
QY 301 GGAGCCACTGCTGTTGGACTTACTTACCCATACCACTATGCTATGCTGATGGGGTGAATTT 360
Db |||||
QY 584 GGAGTCACTGCTGTTGGACTTACTTACCCAACTGGTATGCTGATGGGGTGGAGTT 643
Db |||||
QY 361 CAAGGTCAGGCAAGAGAAACCAAGTAAAGGAAAGCAATCTCCCACTGACCCCGGGACAT 420
Db |||||
QY 644 CAAGATCAGGCAAGAGAAACCAAGTAAAGGAAAGTAACTCTCCCACTCACCCTGGGTACAT 703
Db |||||
QY 421 AGCACCCCTAGCCCTACAAAGGACTAGTTCTCTCTAAACCTACATGAACCTCTCGTACC 480
Db |||||
QY 704 GGCACTCTAGCCCTACAAAGGACTAGATCTCTCAAACTACATGAACCTCTCGTACC 763
Db |||||
QY 481 CATACTGCGCTGTTGAGCTTATTTATACCACTCTCTCGGCTCCATGAGTCTCAGCC 540
Db |||||
QY 764 CATACTGCGCTGTTGAGCTTATTTATACCACTCTCTCGGCTCCATGAGTCTCGCC 823
Db |||||
QY 541 CAAACCCCTACTACTCTTGGATGTGCTCCCTCCCTGCACCTTCAGGCAATCATTTCAATC 600
Db |||||
QY 824 CAAACCCCTACTACTCTTGGATGTGCTCCCTCCCTGAACTTCAGGCAATCATTTCAATC 883
Db |||||
QY 601 CTGTTCTCTGAACTGGAACCACTTCAGACAGAAATTAACACCTCTCGTTTGTAGTA 660
Db |||||
QY 884 CTGTTCTCTGAACTGGAACCACTTCAGACAGAAATTAACACCTCTCGTTTGTAGTA 943
Db |||||
QY 661 GGACCTCTGTTTCCCACTGGAATTAACCACTCTCAAACTCACTGTTGTAATTT 720
Db |||||
QY 944 GGACCTCTGTTTCCCACTGGAATTAACCACTCTCAAACTCACTGTTGTAATTT 1003
Db |||||
QY 721 AGCAATACTATAGACAAACAGCTCCCAATGCAATCAGGTGGGTAAACCTCCCAACGA 780
Db |||||
QY 1004 AGCAATACTATAGACAAACAGCTCCCAATGCAATCAGGTGGGTAAACCTCCCAACGA 1063
Db |||||
QY 781 ATAGTCTGCTTACCTCAGGAAATTTTGTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 840
Db |||||
QY 1064 ATAGTCTGCTTACCTCAGGAAATTTTGTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1123
Db |||||
QY 841 AATGGCTCTTCTAGAACTATGTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAC 900
Db |||||
QY 1124 AATGGCTCTTCTAGAACTATGTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAC 1183
Db |||||
QY 901 ACTGAACAGATTTATACAACTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 960
Db |||||
QY 1184 ACTGAACAGATTTATACAACTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1243
Db |||||
QY 961 CTTCTCTTCTGTTATCAGACGAGGAGTCTAGGACACTAGTACTGTCATTTGGCAGTATC 1020
Db |||||
QY 1244 CTTCTCTTCTGTTATGAGGACGAGGAGTCTAGGACACTAGTACTGTCATTTGGCAGTATC 1303
Db |||||
QY 1021 ACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAAATTAATGTTGATCGAATCGAACAG 1080
Db |||||
QY 1304 ACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAAATTAATGTTGATCGAATCGAACAG 1363
Db |||||
QY 1081 GTCAGTACTCCCTGGTCACTCTGCAAGATCAACTTAACCTCCCTAGCAGCAGTAGTCTCTT 1140
Db |||||
```

```
Db |||||
QY 1364 GTGCGCGACTCCCTGCTCACCCTTGCAGATCAACTTAACCTCCCTAGCAGCAGTAGTCTCTT 1423
Db |||||
QY 1141 CAAATTCGAAGAGCTTTAGACTTGTAAACCGCAAGAGAGGGGAACTGTTTATTTTATTTA 1200
Db |||||
QY 1424 CAAATTCGAAGAGCTTTAGACTTGTAAACCGCTGAAGAGGGGAACTGTTTATTTTATTTA 1483
Db |||||
QY 1201 GGAGAAGAACGCTGTTTATTTATGTTTAAATCAATCCAGAAATGTCACGTGAGAAAGTTAAAGAA 1260
Db |||||
QY 1484 GGGGAAGAAATGCTGTTTATTTATGTTTAAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 1543
Db |||||
QY 1261 ATTGAGATCGAATCAATATAGACAGAGAGCTTCAAAACACGACGCTGGGGCCTC 1320
Db |||||
QY 1544 ATTGAGATCGAATCAATCAACGCTAGACAGAGGAGCTTCGAAACACCTGGACCCCTGGGGCCTC 1603
Db |||||
QY 1321 CTCAGCCCAATGATCCCTGGGTTCTCCCTCTCTTAGGACCTCTAGCAGCTCTAATATTTG 1380
Db |||||
QY 1604 CTCAGCCCAATGATCCCTGGGTTCTCCCTCTCTTAGGACCTCTAGCAGCTCTAATATTTG 1663
Db |||||
QY 1381 TTACTCTCTTTGGACCCCTGATATCTTTAAACCTCTCTGTTTAAAGTTGTTCTCTCCAGAAAT 1440
Db |||||
QY 1664 CTACTCTCTTTGGACCCCTGATATCTTTAAACCTCTCTGTTTAAAGTTGTTCTCTCTCCAGAAAT 1723
Db |||||
QY 1441 GAAGCTGTAAGCTACAGATGCTTACAAATGGGACCCCA 1481
Db |||||
QY 1724 GAAGCTGTAAGCTACAAATGGGACCCCAAGATGCGAGTCCAA 1764
Db |||||

RESULT 26
AF52052982 2694 bp DNA linear PRI 11-PEB-2004
LOCUS Homo sapiens individual 78 allele A, envelope glycoprotein gene,
DEFINITION complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION AF520530
VERSION AF520530.1 GI:33411028
KEYWORDS
SEGMENT
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2694)
AUTHORS Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
Bonnaud,B., Lucotte,G., Duret,B. and Mandrand,B.
TITLE The endogenous retroviral locus ERVWE1 is a bona fide gene involved
in hominoid placental physiology
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
PUBMED 14757826
REFERENCE 2 (bases 1 to 2694)
AUTHORS Mallet,F., Bouton,O. and Oriol,G.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée
d'Italie, Lyon 69364 cedex 07, France
FEATURES
source
1..2694
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="individual 78 allele A"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-q22"
/sex="female"
/cell_type="PBMC's"
/notes="Caucasian
endogenous_virus: HERV-W"
misc_feature 43..57
/notes="splice acceptor site"
284..1900
/notes="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAQ17587.1"
```


FEATURES	d'Italie, Lyon 69364 cedex 07, France	
	Location/Qualifiers	
source	1..2694	
	/organism="Homo sapiens"	
	/mol_type="genomic DNA"	
	/isolate="individual 71 allele A"	
	/db_xref="taxon:9606"	
	/chromosome="7"	
	/map="7q21-q22"	
	/sex="male"	
	/cell_type="PBMC's"	
	/notes="African"	
misc_feature	43..57	
	endogenous_virus: HERV-W"	
CDS	284..1900	
	/note="splice acceptor site"	
	/note="syncytin"	
	/codon_start=1	
	/product="envelope glycoprotein"	
	/protein_id="AA017603.1"	
	/db_xref="GI:33411093"	
	/translations="MALPVIHLFTVLLPSFTLTAPPCCMTSSSPYOEFLLRWORP	
	GNIDARSYSLSGTFTFTAHMPNCHYHSATLCHNANTHYWTKMINPSCEGLGV	
	TVCWTFYFTGMDGGVQDAREKHVKEISQLTRVHGTSFSPYKGLDLSKHELTIR	
	HTRLVSLFTNTLTLGLHSAQNPNCMI CLPLNFRPVYVIPPEQWNESTEINTSV	
	LVGPLVSNLEI HTSNLTCPKFSNTTYTNSQCIRWVTPPTQIVCLPSGIFVFCGTS	
	YRLGSSSESMCFSLVLPWMTIYTDOLYNYVISKPRNKRVPILPFIIVGAGVLGAL	
	TGIGGTTSTQFYKLSQELNGDMERVAADSLVTLQDLSLAIVLQNRALDILTAE	
	RGGTCLPLGECYVYNSQGIYVEKVEIRDLQRAELRWTPGWLISQMPWILP	
	FLGPLAAIILLGLFGPCIFNLNLVFSRRISAEVQLQWEPWQSKTKIYRRLDRPAS	
misc_feature	1705..1720	
	RSVDNIDKGTTPPEIISAQPLLRPSAGSS"	
LTR	1705..1720	
	/note="splice acceptor site"	
ORIGIN	1959..2694	
	Query Match 90.2%; Score 1335.4; DB 9; Length 2694;	
	Best Local Similarity 93.9%; Pred. No. 0;	
	Matches 1390; Conservative 0; Mismatches 91; Indels 0; Gaps 0;	
QY	1 ATGGCCCTCCCTTATCATATCTTTCTTTACTGTTCTCTTACCCCTTTCGCTCTCACT	
	60	
DB	284 ATGGCCCTCCCTTATCATATCTTTCTTTACTGTTCTCTTACCCCTTTCGCTCTCACT	
	343	
QY	61 GCACCCCTCATGCTGCTGTACACAGTAGTCCCTTTACCAAGAGTTTCTATGAAGA	
	120	
DB	344 GCACCCCTCATGCTGCTGTATGACACAGTAGTCCCTTTACCAAGAGTTTCTATGAAGA	
	403	
QY	121 ACGCGGTTCTGGAATATTGATGCCCCCATATATAGGAGTTTATCTAAGGGAACTCC	
	180	
DB	404 ATGCAGGCTCCGGAAATATTGATGCCCATCGTATAGGAGTTTCTAAGGGAACTCC	
	463	
QY	181 ACCTTCACTGCCACACCCATATGCCCGCACTGCTATACCTGCTTTCGCTTTCGATG	
	240	
DB	464 ACCTTCACTGCCACACCCATATGCCCGCACTGCTATACCTGCTTTCGCTTTCGATG	
	523	
QY	241 CATCAATACTCATATTGGACAGGAAATGATTAACTCTAGTTCTCTGGAGGACTT	
	300	
DB	524 CATCAATACTCATATTGGACAGGAAATGATTAACTCTAGTTCTCTGGAGGACTT	
	583	
QY	301 GGAGCCACTGCTGTTGGACTTCTTACCCATACCAAGTATGCTGATGGGGTGGAAAT	
	360	
DB	584 GGAGTCACTGCTGTTGGACTTCTTACCCAACTGCTATGCTGATGGGGTGGAGTT	
	643	
QY	361 CAAGTCAGGCAAGAGAAACAGTAAGAGCAATCTCCCACTGACCCGGGGACAT	
	420	
DB	644 CAAGATCAGGCAAGAGAAACAGTAAGAGCAATCTCCCACTGACCCGGGGACAT	
	703	
QY	421 AGCAACCTTAGCCCTTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCTCCGTAAC	
	480	
DB	704 GGCACCTTAGCCCTTACAAAGGACTAGTTCTCTCAAAACTACATGAAACCTCCGTAAC	
	763	
QY	481 CATACTCGCTGGTGAGCCTATTATTAACCAACCTCACTCGGCTCCATGAGGTCTCAGCC	
	540	

RESULT 28
AX355872
LOCUS
DEFINITION
ACCESSION

AX355872 2930 bp DNA linear PAT 06-FEB-2002
Sequence 1 from Patent WO0204678.
AX355872

1441 GAAAGCTGTTAAAGCTACAGATGGTCTTACAAATGGAAACCCCA 1481
1724 GAAGTGTTAAAGCTACAAATGGAGCCCAAGATGAGTCCAA 1764

764 CATACTCGCTGGTAAAGCTATTATTAACCCCTCACTGGGCTCATGAGGTCTCGGCC 823
541 CAAAAACCTTACTAACTGTTGGATGTGCTTCCCTTGCACCTTTCAGGCCATACATTTCAATC 600
824 CAAAAACCTTACTAACTGTTGGATGTGCTTCCCTTGCACCTTTCAGGCCATATGTTCAATC 883
601 CCTGTTCTCGAAACAATTTGAGCAACAATTTGAGCAACAATAAACACCACTTCCGTTTATGTA 660
884 CCTGTACTGTAACAATTTGAGCAACAATTTGAGCAACAATAAACACCACTTCCGTTTATGTA 943
661 GGACCTCTGTTTCCAAATCTGGAATAACCCATACCTCAAACTCAGCTGTTGTAATAATTT 720
944 GGACCTCTGTTTCCAAATCTGGAATAACCCATACCTCAAACTCAGCTGTTGTAATAATTT 1003
721 AGCAATCTATATAGACACCAACCACTGCTCCCAATGCATCAGTGGGTAAACCTCCCAACCA 780
1004 AGCAATCTATATAGACACCAACCACTGCTCCCAATGCATCAGTGGGTAAACCTCCCAACCA 1063
781 ATAGTCTGCTACCTCCTCAGGAATATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 840
1064 ATAGTCTGCTACCTCCTCAGGAATATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1123
841 AATGGCTCTTCAAGATCTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 900
1124 AATGGCTCTTCAAGATCTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1183
901 ACTGAACAAGATTTATACAAATCATGTCGTACCTAAGCCCCACCAACAAAGAGTACCATT 960
1184 ACTGAACAAGATTTATACAAATCATGTCGTACCTAAGCCCCACCAACAAAGAGTACCATT 1243
961 CTTCTCTTTGTTATACAGACGAGGTGCTAGGACAGTACTAGGCACTGGCACTGGCAGTATC 1020
1244 CTTCTCTTTGTTATAGGAGCAGGAGTGTAGGTGTCATAGGTGTTGTTGTTGTTGTTGTTG 1303
1021 AGAACCTTACTCAGTTCTTACTACAAATCTCTCAAGAAATAAATGGTGCATGGAACAG 1080
1304 ACAACCTTACTCAGTTCTTACTACAAATCTCTCAAGAAATAAATGGTGCATGGAACAG 1363
1081 GTCACCTGACTCCCTGGTCACTTGTGCAAGATCACTTAACTCCCTAGCAGCAGTAGTCCCTT 1140
1364 GTCGCCGACTCCCTGGTCACTTGTGCAAGATCACTTAACTCCCTAGCAGCAGTAGTCCCTT 1423
1141 CAAAATCGAAGAGCTTTAGACTTGTACCGCCCAAGAGAGGGGACCTGTTTATTTTAA 1200
1424 CAAAATCGAAGAGCTTTAGACTTGTACCGCCGTAAGAGAGGGGAAACCTGTTTATTTTAA 1483
1201 GGAGAGAGAGCTGTTTATTTATGTTATCAATCCAGAAATGTCTCACTGAGAAAGTTAAAGAA 1260
1484 GGGAGAGAGAGCTGTTTATTTATGTTATCAATCCAGAAATGTCTCACTGAGAAAGTTAAAGAA 1543
1261 ATTTCGAGATCGAATACAAATGTAGAGCAGAGGAGCTTCAAAAACACCGAACGCTGGGGCCTC 1320
1544 ATTTCGAGATCGAATACAAATGTAGAGCAGAGGAGCTTCAAAAACACCGAACGCTGGGGCCTC 1603
1321 CTGAGCCAAATGAGTGCCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
1604 CTGAGCCAAATGAGTGCCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1663
1381 TTACTCTCTTTGGACCTGTATCTTTAACTCTCTTTAAAGTTTGTCTCTCCAGAAAT 1440
1664 CTACTCTCTTTGGACCTGTATCTTTAACTCTCTTTGTTAACTTTGTTCTCTTCCAGAAATC 1723
1441 GAAAGCTGTTAAAGCTACAGATGGTCTTACAAATGGAAACCCCA 1481
1724 GAAGTGTTAAAGCTACAAATGGAGCCCAAGATGAGTCCAA 1764

VERSION	AX355872.1	GI:18620523
KEYWORDS	Homo sapiens (human)	
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 Keith, J.C., McCoy, J.M. and Mi, S.	
AUTHORS	Methods and compositions for diagnosing and treating preeclampsia	
TITLE	and gestational trophoblast disorders	
JOURNAL	Patent: WO 0204678-A 1 17-JAN-2002;	
GENETICS	INSTITUTE, INC. (US)	
FEATURES	Location/Qualifiers	
source	1..2930	
	/organism="Homo sapiens"	
	/mol_type="unassigned DNA"	
	/db_xref="taxon:9606"	
CDS	930..2546	
	/note="unnamed protein product"	
	/codon_start=1	
	/protein_id="CAD23000.1"	
	/db_xref="GI:18620524"	
	/translations="MALPYHIFLFTVLLPSFTLTAPPKRCMTSSSPYOBFLWMORP GNIDAPSYRLSGTPTFTATHMPNCHVSHATLQMHANTHYWTGKMINPSCPGGLGV TVCWTYVTOGMSDGGVDQAREKHKEVISQLTRVHGTSPPYKGLDLSKLHETLR HTRVLSLNFNLTGLHEVSAQNPNCWICLPLNFRPVSIPVPEQNNFSTEINTSV LVGLVSNLBITNSLTCVKFSNTTYTNSQIRWTPPTQIVCLPSGIFVCGTSA YRLCNGSSMCSLFLVPPMTTYTEQDLYNYVISKPRNRPVILPFLVIGAGVIGALG TGIGGTTSTOFYKLSQELNGDMERVADSLVTLQDLSLAALVFNQRALDLTAE RGCTCLPLGECYVYNSGIVTEKVEIRDRIRRAEELRNTGPGWLLSQMWPILP FLGPLAAILLLLPFCIFNLNVFVSSRIEAVKLQNEPKMQSKTKIYRPLDRASP RSDVNDIKGTPPEISAAQPLLRPNAGSS"	
ORIGIN	90.2%; Score 1335, 4; DB 6; Length 2930;	
Query Match	Best Local Similarity 93.9%; Pred. No. 0;	
Matches 1390; Conservative	0; Mismatches 91; Indels 0; Gaps 0;	
Qy	1	ATGCGCCCTCCCTATCATCTTTCTCTTTACTGTTCTCTTACCCCTTCCTGCTCTCACT 60
Db	930	ATGCGCCCTCCCTATCATATTTTCTTTACTGTTCTTTTACCTTTTACCTCTCTCACT 999
Qy	61	GCACCCCTCATGCTGCTGTACACAGTAGTCTCCCTTACCAGAGTTTCTATGAGA 120
Db	990	GCACCCCTCATGCTGCTGTATGACCATGCTCCCTTACCAGAGTTTCTATGAGA 1049
Qy	121	ACGCGGCTCTCGAATATTGATGCCCATCATATAGGAGTTTATCTAAGGGAATCC 180
Db	1050	ATGAGGGTCCCGGAATATTGATGCCCATCATATAGGAGTTTCTAAGGGAATCC 1109
Qy	181	ACCTTCACTGCCACACCCATATGCCCGCACTGCTATAACTCTGCCACTCTTTGCATG 240
Db	1110	ACCTTCACTGCCACACCCATATGCCCGCACTGCTATCACTCTGCCACTCTTTGCATG 1169
Qy	241	CATGCAATACTCATTTATGGACAGGAAATGATTAATCCTAGTTGCTCTGGAGACTT 300
Db	1170	CATGCAATACTCATTTATGGACAGGAAATGATTAATCCTAGTTGCTCTGGAGACTT 1229
Qy	301	GGAGCACTCTCTGTTGACTTCTTACCCTACCATGATGCTGATGGGGTGGAAAT 360
Db	1230	GGAGTCACTCTCTGTTGACTTCTTACCCTACCATGATGCTGATGGGGTGGAGTT 1289
Qy	361	CAAGGTGAGGCAAGAGAAAACAAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGACAT 420
Db	1290	CAAGATCAGGCAAGAGAAAACATGTAAAGAGTAAATCTCCCACTCACCCGGGTACAT 1349
Qy	421	AGCACCCCTAGCCCTTACAAAGACTAGTTCTCTCAAAAATAATGAAACCCCTCGTACC 480
Db	1350	GGCACCTCTAGCCCTTACAAAGACTAGATCTCTCAAAAATAATGAAACCCCTCGTACC 1409
Qy	481	CATACCTGCTGTGAGCCCTTTTAAATACCACCTCACTCGGCTCCATGAGGTTCTAGCC 540
Db	1410	CATACCTGCTGTGAGCCCTTTTAAATACCACCTCACTCGGCTCCATGAGGTTCTCGGC 1469
Qy	541	CAAAACCCCTACTAACTGTTGGATGTCCTCCCTCCCTGACCTTTCAGGCCATACATTTCATC 600
Db	1470	CAAAACCCCTACTAACTGTTGGATGTCCTCCCTCCCTGAACTTTCAGGCCATATGTTTCAATC 1529
Qy	601	CTGTTTCTTGAAACAATGGAAACAATTCAGCACAGAAATAAACACCACTTCCGTTTGTAGTA 660
Db	1530	CTGTTTCTTGAAACAATGGAAACAATTCAGCACAGAAATAAACACCACTTCCGTTTGTAGTA 1589
Qy	661	GGACCTCTTGTGTTTCCAATCTGGAAATAACCCATACCTCAAACTCTGTTGTAAATTT 720
Db	1590	GGACCTCTTGTGTTTCCAATCTGGAAATAACCCATACCTCAAACTCTGTTGTAAATTT 1649
Qy	721	AGCAATACTATAGACACAACCACTCCCAATGCAATGAGTGGTGAACCTCCACACGA 780
Db	1650	AGCAATACTATACATACACAACCACTCCCAATGCAATGAGTGGTGAACCTCCACACAA 1709
Qy	781	ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGGTACCTCAGCTATCATTTGTTG 840
Db	1710	ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGGTACCTCAGCTATCATTTGTTG 1769
Qy	841	AATGGCTCTTCAGAATCTATGTCTCTCTCATTTCTTAGTGCCCTCATGACCATCTAC 900
Db	1770	AATGGCTCTTCAGAATCTATGTCTCTCTCATTTCTTAGTGCCCTCATGACCATCTAC 1829
Qy	901	ACTGAAACAAGATTATACAATCATGCTGCTACCTAAGCCCCCAACAACAAAGAGTACCATT 960
Db	1830	ACTGAAACAAGATTATACAATCATGCTGCTACCTAAGCCCCCGCAACAAAGAGTACCATT 1889
Qy	961	CTTCTCTTTGTTATCAGACGAGGCTGAGGAGAGTACTAGGCTAGGCTAGGCTAGGCTAGG 1020
Db	1890	CTTCTCTTTGTTATAGGAGGAGGCTGAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGG 1949
Qy	1021	ACAACTCTACTCAGTTCTACTACAAATATCTCAAGAAATAATGCTGACATGGAACAG 1080
Db	1950	ACAACTCTACTCAGTTCTACTACAAATATCTCAAGAAATAATGCTGGAACATGGAACAG 2009
Qy	1081	GTCATGACTCCCTGCTCACCCTTGGCAAGATCAACTTTAACTCCCTAGCAGCAGTAGTCCCT 1140
Db	2010	GTCGCGACTCCCTGCTCACCCTTGGCAAGATCAACTTTAACTCCCTAGCAGCAGTAGTCCCT 2069
Qy	1141	CAAAATCGAAGAGCTTTAGACTTGTAAACCGCCAAAGAGGGGAACTGTTTATTTTA 1200
Db	2070	CAAAATCGAAGAGCTTTAGACTTGTAAACCGCTGAAAGAGGGGAACTGTTTATTTTA 2129
Qy	1201	GGAGAAAGACGCTGTTTATTTATGTTAATCAATCCAGAAATGCTCACTGAGAAAGTTAAAGAA 1260
Db	2130	GGGGAAGATGCTGTTTATTTATGTTAATCAATCCGGAATGCTCACTGAGAAAGTTAAAGAA 2189
Qy	1261	ATTGAGATCGAATACAATGTAGAGCAGAGAGCTTCAAAAACCCGAAACGCTGGGGCTC 1320
Db	2190	ATTGAGATCGAATACAAGGTAGAGCAGAGAGCTTCAAAAACCCGAAACGCTGGGGCTC 2249
Qy	1321	CTCAGCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
Db	2250	CTCAGCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2309
Qy	1381	TTACTCTCTTTGGACCTGTTTCTTTAACTCTCTTTGTTAAAGTTGCTCTCTCCAGAAAT 1440
Db	2310	CTACTCTCTTTGGACCTGTTTCTTTAACTCTCTTTGTTAAAGTTGCTCTCTCCAGAAAT 2369
Qy	1441	GAAGCTGTAAAGTACAGATGCTTTTCAAAATGGAAACCCCA 1481
Db	2370	GAAGCTGTAAAGTACAAATGGAGCCCAAGATGACGTCCAA 2410
RESULT 29	AF208161	
LOCUS	Homo sapiens syncytin precursor, mRNA, complete cds.	
DEFINITION	AF208161	
ACCESSION	AF208161.1	
VERSION	GI:6760400	
KEYWORDS	PRI 22-FEB-2000	

SOURCE	Homo sapiens (human)	1050	ATGCAGCGTCCCGAAATATGTATGCCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC	1109
ORGANISM	Homo sapiens	181	ACCTTCTACTGCCACACCCATATATGCCCGCAACTGCTATTAACCTGCGCATCTTTTGATG	240
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	1110	ACCTTCTACTGCCACACCCATATATGCCCGCAACTGCTATTAACCTGCGCATCTTTTGATG	1169
AUTHORS	1 (bases 1 to 2930) Mi, S.; Lee, X.; Li, X.; Veldman, G. M.; Finnerty, H.; Racine, L.; Lavallie, E.; Tang, X. Y.; Edouard, P.; Howes, S.; Keith, J. C. Jr. and McCoy, J. M.	241	CATGCAAAATCTCATTTATTGGACAGGAAATGATTAATCTAGTCTCTGCGGAGCACTT	300
TITLE	Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis	1170	CATGCAAAATCTCATTTATTGGACAGGAAATGATTAATCTAGTCTCTGCGGAGCACTT	1229
JOURNAL	Nature 403 (6771), 785-789 (2000)	301	GGAGCCACTGCTGTTGGACTTACTTACCCCATACCACTAGTATGCTGATGGGGTGGAAAT	360
MEDLINE	20155476	1230	GGAGTCACTGCTGTTGGACCTTACTTACCCCAACTGCTATGCTGATGGGGTGGAGTT	1289
PUBMED	10693809	361	CAAGTTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT	420
REFERENCE	2 (bases 1 to 2930) Sha, M.; Lee, X.; Li, X.; Veldman, G. M.; Finnerty, H.; Racine, L.; and Lavallie, E.; Tang, X.; Edouard, P.; Howes, S.; Keith, J. C. Jr. and McCoy, J. M.	1290	CAAGATCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGTACAT	1349
AUTHORS	Direct Submission Submitted (26-NOV-1999) Genetics Institute, 87 Cambridge Park Drive, Cambridge, MA 02140, USA	421	AGCACCCCTAGCCCTACAAAGGACTAGTCTCTCAAACTACATGAAACCCCTCCGTACC	480
FEATURES	Location/Qualifiers 1..2930 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="AJ172-28" /tissue_type="testis" 1..2930 /rpt_family="Human endogenous retrovirus HERV-W" 1..529 930..2546 /note="envelope protein" /codon_start=1 /product="syncytin precursor" /protein_id="AAF28334.1" /translation="MALPYHIFLFTVLLPSFTLTAPPCCMTSSPYQFLWRMORP GNIDPVSRLSKGTPPTAHTHPRNCHYSATLCMHANTHYTKMINPSCPGGLV TCHVYQTSQSGGGVQDQAREKHVEVISLTVHGTSSPYKGLDLSKLHETLRT HTRLVSPNTLTGLHEVSAQNPTNCWICLPLNRPVSIIPVPEOWNFSTEINTSV LVGLVSNLEITHSTNLTCVRFSTNTYTTNSQIRWTPPTQIVCLPSGFIFFVCGTSA YRLNGSESMCFKSLFVPPMTIYVEQDLVNVISKPRKRVFIPILFVIGAGVLGALG TGIGITITSTFYKLSQELNGDMERVADELTVLQDLNSLAVALVLRNALDILLTAE RGGLFLFGECCECVNOSGIVTEKVEIKRDIQRRAEELRNTGPMGLLSQWMPWILP FLGPLAAILLLFGPCIFNLLVNFVSRVIAVKLQMEPKMQSKTKIYRPLDRPAP RSDVNDIKGTPPEISAAQPLLRPSAGSS" 930..989 990..2543 /product="syncytin" /note="furin; cleavage site" /evidence=not_experimental 2043..2120 /note="Region: immunosuppressive region" 2235..2357 /note="transmembrane-region site" 2547..2930	1350	GGCACCTCTAGCCCTACAAAGGACTAGATCTCTCAAACTACATGAAACCCCTCCGTACC	1409
		481	CATCTCGCCTGGTGAGCTATTTAATACCAACCCCTCACTCGGCTCCATGAGGTCTCAGCC	540
		1410	CATCTCGCCTGGTGAGCTATTTAATACCAACCCCTCACTCGGCTCCATGAGGTCTCAGCC	1469
		541	CAAAAACCTTAACTGTTGGATGCTCCCTCGCACTTCAAGCCCATATATTCAATC	600
		1470	CAAAAACCTTAACTGTTGGATGCTCCCTCGCACTTCAAGCCCATATATTCAATC	1529
		601	CCTGTTCTCGAACAATGGAAACAACCTTCAAGCAGAGAAATAAACACCACTTCCGTTTGTAGTA	660
		1530	CCTGTACCTGAAACAATGGAAACAACCTTCAAGCAGAGAAATAAACACCACTTCCGTTTGTAGTA	1589
		661	GGACCTCTTGTGTTTCCAATCTGGAATAAACCCATACCTCAAACTCACTCAACCTGTGTAATAATTT	720
		1590	GGACCTCTTGTGTTTCCAATCTGGAATAAACCCATACCTCAAACTCACTCAACCTGTGTAATAATTT	1649
		721	AGCAATATATAGACACACACAGCTCCCAATGCATCAGTGGGTAAACCTCCACACAGA	780
		1650	AGCAATATATAGACACACACAGCTCCCAATGCATCAGTGGGTAAACCTCCACACAGA	1709
		781	ATAGTCTGCTACCCCTCAGGAATATTTTGTCTGTGGTACCTCAGCCCTATCATTTGTTTG	840
		1710	ATAGTCTGCTACCCCTCAGGAATATTTTGTCTGTGGTACCTCAGCCCTATCATTTGTTTG	1769
		841	AATGGCTCTTCAGAAATCTATGTGCTCTCTCATTTCTAGTGGCCCTATGACCATCTAC	900
		1770	AATGGCTCTTCAGAAATCTATGTGCTCTCTCATTTCTAGTGGCCCTATGACCATCTAC	1829
		901	ACTGAAACAAGATTATACAAATCATGTCGTACCTAAGCCCAACAAAGAGAGTACCATT	960
		1830	ACTGAAACAAGATTATACAAATCATGTCGTACCTAAGCCCAACAAAGAGAGTACCATT	1889
		961	CTTCTCTTTTGTATCAGAGCAGGAGTGTAGGCAGACTAGGTACTGGCATTTGGCAGTATC	1020
		1890	CTTCTCTTTTGTATCAGAGCAGGAGTGTAGGTGCTAGGTGCTAGGTACTGGCATTTGGCAGTATC	1949
		1021	ACAACTCTACTCAGTCTTACTACAACTATCTCAAGAAATAAATGCTGACATGGAACAG	1080
		1950	ACAACTCTACTCAGTCTTACTACAACTATCTCAAGAAATAAATGCTGACATGGAACAG	2009
		1081	GTCACTGACCTCCCTGGTCACTTGGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCCT	1140
		2010	GTGCCCCACTCCCTGGTCACTTGGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCCT	2069
		1141	CAAAATCGAAGAGCTTTAGACTTGTCTAAACCGCAAGAGGGGGAACCTGTTTATTTTA	1200
		2070	CAAAATCGAAGAGCTTTAGACTTGTCTAAACCGCTGAAAGAGGGGGAACCTGTTTATTTTA	2129
		1201	GGAGAGAAACGCTGTTTATTTATTAATCAATCCAGAAATCTCACTGAGAAAGTTAAAGAA	1260
		2130	GGGGAAGAATGCTGTTTATTTATTAATCAATCCAGAAATCTCACTGAGAAAGTTAAAGAA	2189

Query Match	90.2%; Score 1335.4; DB 9; Length 2930;
Best Local Similarity	93.9%; Pred. No. 0;
Matches 1390; Conservative	0; Mismatches 91; Indels 0; Gaps 0;
QY	1 ATGGCCCTCCCTATCATCTTTTCTTACTGTTCTTACCCCTTTTGGCTCTCACT 60
DB	930 ATGGCCCTCCCTATCATCTTTTCTTACTGTTCTTACCCCTTTTGGCTCTCACT 989
QY	61 GCACCCCTCCATGCTGCTGTAACACAGTAGTCCCTTACCAAGAGTTCTATGAGA 120
DB	990 GCACCCCTCCATGCTGCTGTAACACAGTAGTCCCTTACCAAGAGTTCTATGAGA 1049
QY	121 ACCGGCTTCTGGAATATGATGCCCATCATATAGGAGTTTATTAAGGGAACCTCC 180

us-09-319-156b-9.rge

Tue Jan 25 08:13:09 2005

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2005, 11:11:14 ; Search time 679.669 Seconds
(without alignments)
11438.492 Million cell updates/sec

Title: US-09-319-156B-9
Perfect score: 1481
Sequence: 1 atggccctcccttatcatac.....gtcttacaatggaacccca 1481

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04.*

1: Geneseqn1980s.*
2: Geneseqn1930s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1481	100.0	1481	2	AAV43217 Multiple
2	1481	100.0	1481	2	AAV43217 Multiple
3	1481	100.0	1481	12	ADG14846
4	1481	100.0	2030	3	AAA63826 Nucleotid
5	1461.8	98.7	1629	3	AAA96625 DNA encod
6	1450	97.9	1481	9	ADB84400 MSRV-1 as
7	1389.8	93.8	2074	6	AD41225 Human EMB
8	1335.4	90.2	2930	6	AD24195 Human syn
9	1335.4	90.2	2946	2	AAZ77526 Human sec
10	1335.4	90.2	2946	3	AAZ59468 Human sec
11	1335.4	90.2	2946	10	ADC38776 Human cDN
12	1333.8	90.1	1617	5	AAH20070 HERV-W en
13	1333.8	90.1	2055	3	ABN97948 Human ret
14	1333.8	90.1	2599	3	ABN97927 Human ret
15	1333.8	90.1	10499	3	ABN97929 Human ret
16	1333.8	90.1	56093	6	ABL61744 Colon ade
17	1330.6	89.8	2781	5	AAF55630 Nucleotid
18	1324.2	89.4	7582	3	AAV59215 Human end
19	1322.6	89.3	7582	2	AAV25665 Complete
20	1308.2	88.3	2782	5	AAH20069 HERV-W en
21	1306.6	88.2	2782	2	AAV25661 Human end

22	1306.6	88.2	2782	3	AAV59211
23	1298.2	87.7	3464	10	AD509587
24	1298.2	87.7	9502	10	AD59718
25	1282	86.6	46340	3	ABN97978
26	1278.8	86.3	1799	3	ABN97931
27	1239	83.7	2784	3	ABN97930
28	1131.6	76.4	1894	4	AAI14608
29	1131.6	76.4	1894	4	ABA56337
30	1131.6	76.4	1894	4	AAI35980
31	1131.6	76.4	1894	4	ABA45822
32	1131.6	76.4	1894	4	ABA25978
33	1131.6	76.4	1894	4	AAK30018
34	1131.6	76.4	1894	4	AAK04516
35	1131.6	76.4	1894	4	ABS29670
36	1131.6	76.4	1894	5	AAI04422
37	1131.6	76.4	1894	6	ABS04589
38	1116.4	75.4	1948	2	AAV25659
39	1116.4	75.4	1948	3	AAV59209
40	1048	70.8	6394	5	AAV84210
41	716.6	48.4	792	4	AAI23803
42	716.6	48.4	792	4	ABA68919
43	716.6	48.4	792	4	AAI49115
44	716.6	48.4	792	4	ABA50940
45	716.6	48.4	792	4	ABA35872

ALIGNMENTS

RESULT 1
AAV43217
ID AAV43217 standard; cDNA; 1481 BP.
XX
AC AAV43217;
XX
DT 29-DEC-1998 (first entry)
XX
DE Multiple sclerosis associated retrovirus fragment 5.
XX
KW Multiple sclerosis associated retrovirus; MSRV; MS; pol gene; gag gene;
KW env gene; rheumatoid arthritis-associated virus; ss.
XX
OS Multiple sclerosis associated retrovirus.
XX
FH Location/Qualifiers
FT CDS
FT 1..1479
FT /*tag= a
FT /product= "Encodes protein AAV71068"
FT /transl_except= (pos:115-117, appears to code for a stop
FT codon)
FT /note= "CDS does not contain a stop codon"
XX
PN WO9823755-A1.
XX
PD 04-JUN-1998.
XX
PF 26-NOV-1997; 97WO-IB001482.
XX
PR 26-NOV-1996; 96US-00756429.
XX
PA (INMR) BIO MERIEUX.
XX
PI Perron H, Beseme F, Bedin F, Paranhos-Baccala G;
PI Komurian-Pradel F, Jollivet-Reynaud C, Mandrand B;
XX
DR WPI; 1998-322732/28.
XX
DR P-PSDB; AAV71068.
XX
PT New nucleic acid from retroviruses - useful for diagnosis, prevention and
PT treatment of, e.g. multiple sclerosis.
XX
PS Disclosure; Page 184-185; 286pp; English.
XX

XX Nucleic acid sequences of retrovirus called MSRV-1 - associated with
PT multiple sclerosis or rheumatoid polyarthritis.
XX
PS Claim 1; Page 37-38; 83pp; French.
XX
CC This sequence represents clone C15 from a novel multiple sclerosis
CC related virus type 1 (MSRV1). The sequence can be used in diagnostic,
CC prophylactic or therapeutic compositions to inhibit expression of a
CC multiple sclerosis related virus and/or virus associated with rheumatoid
CC polyarthritis. (Updated on 27-AUG-2003 to correct OS field.) (Updated on
CC 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 1481 BP; 412 A; 410 C; 261 G; 398 T; 0 U; 0 Other;

Query Match 100.0%; Score 1481; DB 2; Length 1481;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCCCTCCCTATCATACTTTCTCTTTACTGTTCTCTTACCCGCTTTCGCTCTCACT 60
DB 1 ATGGCCCTCCCTATCATACTTTCTCTTTACTGTTCTCTTACCCGCTTTCGCTCTCACT 60

QY 61 GCACCCCTCCATGCTGCTCAACACGAGTCTCCCTTACCAGAGTTTCTATGAAGA 120
DB 61 GCACCCCTCCATGCTGCTCAACACGAGTCTCCCTTACCAGAGTTTCTATGAAGA 120

QY 121 ACAGCGCTTCTCTGGAATATTGATGCCCCCATATAGAGGTTTATCTAAGGGAATCTC 180
DB 121 ACAGCGCTTCTCTGGAATATTGATGCCCCCATATAGAGGTTTATCTAAGGGAATCTC 180

QY 181 ACCTTCACTGCCCCACACCCATATGCCCCGCACTGCTATTAACCTCTGCGACTCTTTCGATG 240
DB 181 ACCTTCACTGCCCCACACCCATATGCCCCGCACTGCTATTAACCTCTGCGACTCTTTCGATG 240

QY 241 CATGCAAACTACTATTATGAGCAGGGAATGATTAATCTAGTTGTCCTGAGGACTT 300
DB 241 CATGCAAACTACTATTATGAGCAGGGAATGATTAATCTAGTTGTCCTGAGGACTT 300

QY 301 GGAGCCACTCTCTGTTGGACTTACTTCAACCACCATACAGTATGCTGATGGGGTGGAAAT 360
DB 301 GGAGCCACTCTCTGTTGGACTTACTTCAACCACCATACAGTATGCTGATGGGGTGGAAAT 360

QY 361 CAAGGTGAGCAAGAGAAAACAAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGACAT 420
DB 361 CAAGGTGAGCAAGAGAAAACAAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGACAT 420

QY 421 AGCACCCTAGCCCTACAAAGCACTAGTTCTCTCAAACTACATGAACCTTCCGTACC 480
DB 421 AGCACCCTAGCCCTACAAAGCACTAGTTCTCTCAAACTACATGAACCTTCCGTACC 480

QY 481 CATACTCGCTGCTGAGCCCTATTATTAATACCACTCCTCACTCGGCTCCATGAGGTCTCAGCC 540
DB 481 CATACTCGCTGCTGAGCCCTATTATTAATACCACTCCTCACTCGGCTCCATGAGGTCTCAGCC 540

QY 541 CAAAACCTACTAACTGTTGGATGTCCTCCCTCGCACTTCAGGCCATACATTTCAATC 600
DB 541 CAAAACCTACTAACTGTTGGATGTCCTCCCTCGCACTTCAGGCCATACATTTCAATC 600

QY 601 CTTGTTCTGCAACAAATGGAACTTTCAGCAGCAAGAAATAAACACCACTTCCGTTTATGA 660
DB 601 CTTGTTCTGCAACAAATGGAACTTTCAGCAGCAAGAAATAAACACCACTTCCGTTTATGA 660

QY 661 GGACCTCTGTTTCCAACTGGAATAAACCCATACCTCAAACTCCTGCTGTGTAATAATTT 720
DB 661 GGACCTCTGTTTCCAACTGGAATAAACCCATACCTCAAACTCCTGCTGTGTAATAATTT 720

QY 721 AGCAATACTATAGACAAACACGCTCCCAATGATCAGGTGGGTAAACCTCCACACGA 780
DB 721 AGCAATACTATAGACAAACACGCTCCCAATGATCAGGTGGGTAAACCTCCACACGA 780

QY 781 ATAGTCTGCTACCTCAGGAATATTTTGTGTTGTTACCTCAGGCTATCATTTGTTG 840
DB 781 ATAGTCTGCTACCTCAGGAATATTTTGTGTTGTTACCTCAGGCTATCATTTGTTG 840

QY 841 AATGGCTCTTCAAGATCTATGTCCTCTCATTTCTAGTCCCTTATGACCATCTAC 900
DB 841 AATGGCTCTTCAAGATCTATGTCCTCTCATTTCTAGTCCCTTATGACCATCTAC 900

QY 901 ACTGAACAGATTTATACAACTATGTCGTCACCTTAAGCCCCACAAAGAGTACCCATT 960
DB 901 ACTGAACAGATTTATACAACTATGTCGTCACCTTAAGCCCCACAAAGAGTACCCATT 960

QY 961 CTTCTCTTTGTTTATCAGACGAGAGTGTAGGAGACTAGGTACTGGCATTGGCAGTATC 1020
DB 961 CTTCTCTTTGTTTATCAGACGAGAGTGTAGGAGACTAGGTACTGGCATTGGCAGTATC 1020

QY 1021 ACAACCTCTACTCAGTTCTTACTACAACTATCTCAAGAAATAAATGCTGACATGGAACAG 1080
DB 1021 ACAACCTCTACTCAGTTCTTACTACAACTATCTCAAGAAATAAATGCTGACATGGAACAG 1080

QY 1081 GTCACTGACTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
DB 1081 GTCACTGACTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140

QY 1141 CAAAATCGAAGAGCTTTAGACTTTGCTAAACCGCCAAAGAGGGGGAACCTGTTTATTTTAA 1200
DB 1141 CAAAATCGAAGAGCTTTAGACTTTGCTAAACCGCCAAAGAGGGGGAACCTGTTTATTTTAA 1200

QY 1201 GGAGAAAGACGCTGTTTATTTATGTTATTAATCAATCCAGAAATGCTCACTGAGAAAGTTAAAGAA 1260
DB 1201 GGAGAAAGACGCTGTTTATTTATGTTATTAATCAATCCAGAAATGCTCACTGAGAAAGTTAAAGAA 1260

QY 1261 ATTGAGATCGAATACAAATGTTAGAGGAGGAGCTTCAAAACACCGCCGCTGCGGCTC 1320
DB 1261 ATTGAGATCGAATACAAATGTTAGAGGAGGAGCTTCAAAACACCGCCGCTGCGGCTC 1320

QY 1321 CTCAGCAATGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
DB 1321 CTCAGCAATGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380

QY 1381 TTACTCTCTTTGGACCCCTGATCTTTAACTCTTAACTCTGTTAAAGTTTCTCTTCCAGAAAT 1440
DB 1381 TTACTCTCTTTGGACCCCTGATCTTTAACTCTTAACTCTGTTAAAGTTTCTCTTCCAGAAAT 1440

QY 1441 GAAGCTGTTAAAGTACAGATGCTTACAAATGGAAACCCCA 1481
DB 1441 GAAGCTGTTAAAGTACAGATGCTTACAAATGGAAACCCCA 1481

RESULT 3
ADG14846
ID ADG14846 standard; cDNA; 1481 BP.
XX
XX ADG14846;
XX AC
XX XX
DT 26-FEB-2004 (first entry)
XX
DE MSRV associated cDNA #9.
XX
XX KW
XX XX
XX ss; pol gene; retrovirus; multiple sclerosis; rheumatoid arthritis.
XX Multiple sclerosis associated retrovirus.
OS
XX US2003198647-A1.
PN
XX 23-OCT-2003.
PD
XX 03-APR-2002; 2002US-00114104.
PF
XX 26-NOV-1996; 96US-00756429.
PR
XX 26-NOV-1997; 97US-00979847.
PR
XX (INMR) BIO MERIEUX.
PA
XX Perron H, Beseme F, Bedin F, Paranhos-Baccala G;
PI


```
FT CAAT_signal 1800..1807
FT FT /*tag= C
FT CAAT_signal 1858..1864
FT FT /*tag= d
FT TATA_signal 1906..1911
FT FT /*tag= e
FT polyA_signal 1996..2002
FT FT /*tag= f
XX
XX WO200047745-A1.
XX
XX 17-AUG-2000.
XX
XX 15-FEB-2000; 2000WO-IB000159.
XX
XX 15-FEB-1999; 99EP-00420041.
XX
XX (INMR ) BIO MERIEUX.
XX
XX Paranhos-Baccala G, Perron H, Komurian-Pradel F;
XX
XX WPI; 2000-506097/45.
XX
XX P-PSDB; AAB08195.
XX
XX Nucleotide fragment of LTR-RUS region from Multiple Sclerosis retrovirus
XX (MSRV) used to detect the presence of MSRV-1 retrovirus in a biological
XX sample.
XX
XX Disclosure; Fig 2; 23pp; English.
XX
XX The present sequence represents the nucleotide sequence corresponding to
XX the 3' env region and long terminal repeat sequences from clone C16 of
XX Multiple Sclerosis retrovirus (MSRV-1). The specification describes a
XX long terminal repeat (LTR)-RUS region which encodes the expression of a
XX MSRV-1 protein. This is unusual for LTRs, in particular in the RUS
XX region. The sequence includes CAAT and TATA signals which are present in
XX the U3 and R regions and are not directed towards the CDS indicated in
XX the features table. Probes and antibodies to the MSRV-1 retrovirus
XX protein and encoding polynucleotide sequences are used to detect the
XX presence of MSRV-1 retrovirus in a biological sample. (Updated on 06-AUG-
XX 2003 to correct OS field.)
XX
XX Sequence 2030 BP; 574 A; 559 C; 387 G; 510 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 1481; DB 3; Length 2030;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 ATGGCCCTCCCTATCATACTTTTCTTTTACTGTTCTCTTACCCCTTTGCTCTCACT 60
XX 1 ATGGCCCTCCCTATCATACTTTTCTTTTACTGTTCTCTTACCCCTTTGCTCTCACT 60
XX
XX 61 GCACCCCTCATGCTCTGTACACAGTAGTCTCCCTTACCAGAGTTTCTATGAGA 120
XX 61 GCACCCCTCATGCTCTGTACACAGTAGTCTCCCTTACCAGAGTTTCTATGAGA 120
XX
XX 121 ACGGGGCTCTCTGAAATATTGATGCCCATCATATAGGAGTTTATCTAAGGAAACTCC 180
XX 121 ACGGGGCTCTCTGAAATATTGATGCCCATCATATAGGAGTTTATCTAAGGAAACTCC 180
XX
XX 181 ACCTTCACTGCCACACCCATATGCCCGCAACTGCTATAACTTGCCACTCTTTGCATG 240
XX 181 ACCTTCACTGCCACACCCATATGCCCGCAACTGCTATAACTTGCCACTCTTTGCATG 240
XX
XX 241 CATGCAATACTCATTTATGGAAGGAAATGATTAATCTAGTTGCTCTGAGGACTT 300
XX 241 CATGCAATACTCATTTATGGAAGGAAATGATTAATCTAGTTGCTCTGAGGACTT 300
XX
XX 301 GGAGCCACTGCTCTGGACTTACTTCCACCATACCAAGTATGCTGATGGGGTGAATT 360
XX 301 GGAGCCACTGCTCTGGACTTACTTCCACCATACCAAGTATGCTGATGGGGTGAATT 360
XX
XX 361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGACAT 420
XX
DB 361 CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGACAT 420
QY 421 AGCACCCCTAGCCCTTACAAAGGACTAGTTTCTTCAAAACTACATGAAACCTTCGTTACC 480
DB 421 AGCACCCCTAGCCCTTACAAAGGACTAGTTTCTTCAAAACTACATGAAACCTTCGTTACC 480
QY 481 CATACTCGCTGCTGAGCCCTTATTAATACCAACCTCCTCGGCTCCATGAGGTTCTAGCC 540
DB 481 CATACTCGCTGCTGAGCCCTTATTAATACCAACCTCCTCGGCTCCATGAGGTTCTAGCC 540
QY 541 CAAAACCTTACTAACTGTTGGATGTGCTCCCTCCCTGCACTTCAGGCCATACATTTCAATC 600
DB 541 CAAAACCTTACTAACTGTTGGATGTGCTCCCTCCCTGCACTTCAGGCCATACATTTCAATC 600
QY 601 CTTGTTCTTGAAACAATGGAAACAATTTTACGACAGAAATAAACACCACTTCCGTTTGTAGTA 660
DB 601 CTTGTTCTTGAAACAATGGAAACAATTTTACGACAGAAATAAACACCACTTCCGTTTGTAGTA 660
QY 661 GGACCTCTTGTGTTTCCAAATCTGGAAATAACCCATACCTCAAACTCCTGTTGTAATAATTT 720
DB 661 GGACCTCTTGTGTTTCCAAATCTGGAAATAACCCATACCTCAAACTCCTGTTGTAATAATTT 720
QY 721 AGCAATACTATAGACACAACAGCTCCCAATGCATCAGGTGGGTAAACACCTCCACACAGA 780
DB 721 AGCAATACTATAGACACAACAGCTCCCAATGCATCAGGTGGGTAAACACCTCCACACAGA 780
QY 781 ATAGTCTGCTTACCTCAGGAATATTTTGTCTGTGGTACCTCAGCTATCATTTGTTG 840
DB 781 ATAGTCTGCTTACCTCAGGAATATTTTGTCTGTGGTACCTCAGCTATCATTTGTTG 840
QY 841 AATGGCTCTTCAAGATCTATGCTCTCTCTCATTTCTAGTGCCCTTATGACATCTAC 900
DB 841 AATGGCTCTTCAAGATCTATGCTCTCTCTCATTTCTAGTGCCCTTATGACATCTAC 900
QY 901 ACTGAAACAGATTTTATACATCATGCTGCTACCTTAAGCCCAACAAAGAGTACCCATT 960
DB 901 ACTGAAACAGATTTTATACATCATGCTGCTACCTTAAGCCCAACAAAGAGTACCCATT 960
QY 961 CTTCTCTTTGTTATCAGAGCAGGAGTGTAGGAGAGTACTAGGTACTGGCATTGGCAGTATC 1020
DB 961 CTTCTCTTTGTTATCAGAGCAGGAGTGTAGGAGAGTACTAGGTACTGGCATTGGCAGTATC 1020
QY 1021 ACAACCTTACTCAGTTTCTACTACAAACTATCTCAAGAAATAAATGGTGAAGAAACAG 1080
DB 1021 ACAACCTTACTCAGTTTCTACTACAAACTATCTCAAGAAATAAATGGTGAAGAAACAG 1080
QY 1081 GTCACTGACTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
DB 1081 GTCACTGACTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
QY 1141 CAAAATCGAAGAGCTTTAGACTTTGCTAAACCGCAAGAGGGGGAACCTGTTATTTTAA 1200
DB 1141 CAAAATCGAAGAGCTTTAGACTTTGCTAAACCGCAAGAGGGGGAACCTGTTATTTTAA 1200
QY 1201 GGAGAGAGAGCTGTTTATTAATCAATCCAGAAATGCTCACTGAGAAAGTTAAAGAA 1260
DB 1201 GGAGAGAGAGCTGTTTATTAATCAATCCAGAAATGCTCACTGAGAAAGTTAAAGAA 1260
QY 1261 ATTGAGATCGAATACAAATGTAGACAGAGAGCTTCAAAACACCGAACCTGGGGCTC 1320
DB 1261 ATTGAGATCGAATACAAATGTAGACAGAGAGCTTCAAAACACCGAACCTGGGGCTC 1320
QY 1321 CTGAGCAATGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
DB 1321 CTGAGCAATGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
QY 1381 TTACTCTCTTTGGACCCCTGTATCTTTAACTCTTGTGTTAGTTTGTCTCTCCAGAAAT 1440
DB 1381 TTACTCTCTTTGGACCCCTGTATCTTTAACTCTTGTGTTAGTTTGTCTCTCCAGAAAT 1440
QY 1441 GAAGGCTGTAAGGTACAGATGGTCTTACAAATGGAACCCCA 1481
DB 1441 GAAGGCTGTAAGGTACAGATGGTCTTACAAATGGAACCCCA 1481
```

Db	1441	GAAGCTGTTAAAGCTACAGATGGTCTTTACAAATGGAACCCCA	1481
Db	121	ACGCGGCTTCTCGGAAATATTTGATGCCCCCATCATATAGGAGTTTATCTAAGGGAAATCC	180
Db	181	ACCTTCACTGCCACACCCATATGCCCCGCAACTGCTATTAACCTCTGCACCTCTTTGCAATG	240
Db	181	ACCTTCACTGCCACACCCATATGCCCCGCAACTGCTATTAACCTCTGCACCTCTTTGCAATG	240
Qy	241	CATGCAAACTACTCATTATTGGACAGGGAATAATCCTAGTCTCTCTGGAGGACTT	300
Db	241	CATGCAAACTACTCATTATTGGACAGGGAATAATCCTAGTCTCTCTGGAGGACTT	300
Qy	301	GGAGCCACTGCTGTGTGGACTTATCTTACCCATACCAAGTATGCTGATGGGGTGGAAAT	360
Db	301	GGAGCCACTGCTGTGTGGACTTATCTTACCCATACCAAGTATGCTGATGGGGTGGAAAT	360
Qy	361	CAAGGTGAGGCAAGAGAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT	420
Db	361	CAAGGTGAGGCAAGAGAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGGACAT	420
Qy	421	AGCACCCCTAGCCCCCTACAAAGGACTAGTCTCTCAAAAATACATAGAAACCTCCGTAAC	480
Db	421	AGCACCCCTAGCCCCCTACAAAGGACTAGTCTCTCAAAAATACATAGAAACCTCCGTAAC	480
Qy	481	CATATCTGGCTGGTGAGCCTATTTAATAACACCCCTCACTGGCTCCATGAGGTCTAGCC	540
Db	481	CATATCTGGCTGGTGAGCCTATTTAATAACACCCCTCACTGGCTCCATGAGGTCTAGCC	540
Qy	541	CAAAACCCCTACTTAAGTGTGGATGCTCCCTCCCTGCACCTTCAGGCGCATACATTTCAATC	600
Db	541	CAAAACCCCTACTTAAGTGTGGATGCTCCCTCCCTGCACCTTCAGGCGCATACATTTCAATC	600
Qy	601	CCTGTTCTCGAACCAATGAAACAACTTCAGCACAGAAAATAAACACCACTTCGTTTGTAGTA	660
Db	601	CCTGTTCTCGAACCAATGAAACAACTTCAGCACAGAAAATAAACACCACTTCGTTTGTAGTA	660
Qy	661	GGACCTCTGTTTCCAAATCTGGAATAATACCCATACCTCAAAACCTCACTGTGTAAATTT	720
Db	661	GGACCTCTGTTTCCAAATCTGGAATAATACCCATACCTCAAAACCTCACTGTGTAAATTT	720
Qy	721	AGCAATATCTATGACACACACAGCTCCCAATGCATCAGGTGGGTAAACACCTCCACACGA	780
Db	721	AGCAATATCTATGACACACACAGCTCCCAATGCATCAGGTGGGTAAACACCTCCACACGA	780
Qy	781	ATAGTCTGCCTACCTCAGGAATATTTTTGCTGTGGTACCTCAGGCTATCATTTGTTTG	840
Db	781	ATAGTCTGCCTACCTCAGGAATATTTTTGCTGTGGTACCTCAGGCTATCATTTGTTTG	840
Qy	841	AATGCTCTTCAGAACTATATGTCTCTCTCATTTAGTGGCCCCCTATGACATCTAC	900
Db	841	AATGCTCTTCAGAACTATATGTCTCTCTCATTTAGTGGCCCCCTATGACATCTAC	900
Qy	901	ACTGAAACAGATTTTATACAATCATGTCTGCTAAAGCCCAACAAAGAGTACCCATT	960
Db	901	ACTGAAACAGATTTTATACAATCATGTCTGCTAAAGCCCAACAAAGAGTACCCATT	960
Qy	961	CTTCTCTTTGTTTATCAGAGCAGGAGTGTCTAGGACAGTCTAGTACTGGCATTTGGCAGTATC	1020
Db	961	CTTCTCTTTGTTTATCAGAGCAGGAGTGTCTAGGACAGTCTAGTACTGGCATTTGGCAGTATC	1020
Qy	1021	ACAACTCTACTCAGTCTTACTACAAACTATCTCAAGAAATAAATGTGTGATGGAACAG	1080
Db	1021	ACAACTCTACTCAGTCTTACTACAAACTATCTCAAGAAATAAATGTGTGATGGAACAG	1080
Qy	1081	GTCACTGTACTCCCTGGTCACTCTGCAAGATCAACTTAACCTCCCTAGCAGCAGTAGTCCCT	1140
Db	1081	GTCACTGTACTCCCTGGTCACTCTGCAAGATCAACTTAACCTCCCTAGCAGCAGTAGTCCCT	1140
Qy	1141	CAAAATCGAAGAGCTTTTACACTGTCTAACCCGCAAAAGAGGGGAAACCTGTTATTTTTA	1200
Db	1141	CAAAATCGAAGAGCTTTTACACTGTCTAACCCGCAAAAGAGGGGAAACCTGTTATTTTTA	1200
Qy	1201	GGAGAAAGACGCTGTTATTTATGTTTAAATCAATTCAGAAATGTCATGAGAAAGTTAAAGAA	1260

1441 GAAGCTGTTAAAGCTACAGATGGTCTTTACAAATGGAACCCCA 1481

RESULT 5

AAA96625

AAA96625 standard; DNA; 1629 BP.

AAA96625;

08-FEB-2001 (first entry)

DNA encoding an envelope (env) protein of MSRV-1.

Envelope protein; MSRV-1; superantigen; autoimmune disease; Vbeta16;

Vbeta17; multiple sclerosis; vaccine; ss.

Human endogenous retrovirus.

Key Location/Qualifiers

CDS 1..1629

FT /*tag= a

FT /product= "envelope protein"

XX

PN W0200057185-A1.

XX

PD 28-SEP-2000.

XX

PF 20-MAR-2000; 2000WO-FR000691.

XX

PR 19-MAR-1999; 99FR-00003622.

XX

PR 28-OCT-1999; 99FR-00013755.

XX

PA (INMR) BIO MERIEUX.

XX

PI Perron H, Lafont M;

XX

DR WPI; 2000-638279/61.

DR

P-PSDB; AAB19069.

XX

Detecting superantigen activity, useful for identifying agents for

treatment or prevention of autoimmune disease, from expansion or loss of

particular lymphocyte Vbeta determinants.

XX

Claim 25; Page 123-124; 134pp; French.

XX

CC The present sequence encodes an envelope protein of MSRV-1. The envelope

protein expressed by the endogenous human retrovirus MSRV-1 has

superantigen activity associated with autoimmune disease. The protein can

be detected using the method of the invention. The specification

CC describes a process for detecting activity of a superantigen in a

biological sample. The process comprises identifying large scale

expansion or loss of lymphocytes that carry at least one of the Vbeta16

and/or Vbeta17 determinants. The method is used to screen for agents that

inhibit the superantigen, especially those associated with MSRV-1 which

is implicated in autoimmune disease, particularly multiple sclerosis.

CC These agents are potentially useful for treatment or prevention (e.g. as

CC vaccines) of autoimmune diseases

XX

SQ Sequence 1629 BP; 456 A; 454 C; 287 G; 432 T; 0 U; 0 Other;

Query Match 98.7%; Score 1461.8; DB 3; Length 1629;

Best Local Similarity 99.2%; Pred. No. 0;

Matches 1469; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 ATGGCCCTCCCTTATCATATCTTTCTCTTTACTGTTCTCTTACCCCTTTCCGCTCTCACT 60

Db 1 ATGGCCCTCCCTTATCATATCTTTCTCTTTACTGTTCTCTTACCCCTTTCCGCTCTCACT 60

Qy 61 GCACCCCTCCATGCTGCTGACACCAAGTACCTCCCTTACCAAGAGTCTTATGAGA 120

Db 61 GCACCCCTCCATGCTGCTGACACCAAGTACCTCCCTTACCAAGAGTCTTATGAGA 120

Qy 121 ACGCGGCTTCTCGGAAATATTTGATGCCCCCATCATATAGGAGTTTATCTAAGGGAAATCC 180

Db 1201 GGAGAGACGCTGTTATTATGTTAATCAATCCAGAAATGCTCACTGAGAAAGTTAAAGAA 1260
QY 1261 ATTGAGATCGAATACAAATGTAGACGAGAGGCTTCAAAACACCGAACCTGGGGCTTC 1320
Db 1261 ATTGAGATCGAATACAAATGTAGACGAGAGGCTTCAAAACACCGAACCTGGGGCTTC 1320
QY 1321 CTCAGCAATGGATGCCCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
Db 1321 CTCAGCCATGGATGGCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATT 1380
QY 1381 TTACTCTCTTTGGACCTGTATCTTTAACTCTCTTGTAAAGTTTGTCTCTCCAGAAAT 1440
Db 1381 TTACTCTCTTTGGACCTGTATCTTCACTCTCTTGTAAAGTTTGTCTCTCCAGAAAT 1440
QY 1441 GAAGCTGTAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
Db 1441 GAAGCTGTAAGCTACAAATAGTTCTTCAATGGAACCCCA 1481

RESULT 6

ADB84400

ID ADB84400 standard; DNA; 1481 BP.

XX ADB84400;

DT 04-DEC-2003 (first entry)

XX MSRV-1 associated DNA sequence #19.

DE ds: multiple sclerosis; rheumatoid arthritis; gag; pol;

KW reverse transcriptase; ribonuclease H.

XX Unidentified;.

OS US2003039664-A1.

PN 27-FEB-2003.

XX 26-NOV-1997; 97US-00979847.

PR 26-NOV-1996; 96US-00756429.

XX (PERR/) PERRON H.

PA (BESE/) BESEME F.

PA (BED/) BEDIN F.

PA (PARA/) PARANHOS-BACCALA G.

PA (KOMU/) KOMURIAN-PRADEL F.

PA (JOLI/) JOLIVET-REYNAUD C.

PA (MAND/) MANDRAND B.

PA (GARS/) GARSON J A.

PA (TUKS/) TUKS P W.

XX Perron H, Beseme F, Bedin F, Paranhos-Baccala G;

PI Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B, Garson JA, Tuke PW;

XX WPI; 2003-512253/48.

DR New isolated or purified nucleic acid associated with multiple sclerosis

XX and/or rheumatoid arthritis, useful for detecting a virus associated with

PT multiple sclerosis or rheumatoid arthritis in a biological sample.

XX Claim 31; Page 77-78; 193pp; English.

PS The invention relates to an isolated or purified nucleic acid from a

XX virus associated with multiple sclerosis and/or rheumatoid arthritis,

CC multiple sclerosis-associated virus (MSRV)-1. The nucleic acids comprise

CC pol, gag or reverse transcriptase genes (or their fragments) encoding the

CC proteins or defined peptides (including immunodominant peptides),

CC antigenic peptides or conserved motifs). Also included are a process for

CC detecting a virus associated with multiple sclerosis or rheumatoid

CC arthritis in a biological sample, a nucleic acid probe for the detection

CC of a virus associated with multiple sclerosis or rheumatoid arthritis, a

CC primer for the amplification by polymerisation of a nucleic acid of a

CC

CC

CC

CC

CC viral material associated with multiple sclerosis or rheumatoid
CC arthritis, a polypeptide exhibiting an inhibitory activity on the
CC proteolytic, reverse transcriptase or ribonuclease H activity from MSRV,
CC and an antibody directed against the MSRV-1 virus obtained by
CC immunologically reacting a human or animal body or cells with an
CC immunogenic agent consisting of the antigenic polypeptide defined above.
CC The nucleic acids are useful for detecting a biological sample a virus
CC associated with multiple sclerosis or rheumatoid arthritis, or for
CC detecting in a biological sample, the presence of or exposure to a virus
CC associated with multiple sclerosis or rheumatoid arthritis. The present
CC sequence is a claimed MSRV-associated sequence whose identity cannot be
CC accurately determined. Note: The SEQ ID numbers for the sequences as
CC displayed in the main body of the patent do not match the SEQ ID numbers
CC in the sequence listing. Consequently those sequences mentioned in the
CC claims may not be the sequences the authors intended to claim.

XX

SQ Sequence 1481 BP; 405 A; 403 C; 257 G; 385 T; 0 U; 31 Other;

Query Match 97.9%; Score 1450; DB 9; Length 1481;

Best Local Similarity 97.9%; Pred. No. 0;

Matches 1450; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1 ATGGCCCTCCCTTATCATACTTTCTTTACGTGTTCTTTACCCCTTTTCGCTCTCACT 60

Db 1 ATGGCCCTCCCTTATCATACTTTCTTTACGTGTTCTTTACCCCTTTTCGCTCTCACT 60

QY 61 GCACCCCTCCATGCTGTACCAACAGTAGTCTCCCTTTACCAAGAGTTTCTATGAAGA 120

Db 61 GCACCCCTCCATGCTGTACCAACAGTAGTCTCCCTTTACCAAGAGTTTCTATGAAGA 120

QY 121 ACGCGCTTCTCGGAAATATTGATGCCCATCATATAGGAGTTTATTAAGGAAACTCC 180

Db 121 ACGCGCTTCTCGGAAATATTGATGCCCATCATATAGGAGTTTATTAAGGAAACTCN 180

QY 181 ACCTTCACTGCCACACCCATATGCCCCGCACTGCTATAAATCTGCACTCTTTGGCATG 240

Db 181 ACCTTCACTGCCACACCCATATGCCCCGCACTGCTATAAATCTGCACTCTTTGGCATN 240

QY 241 CATGCAAAATACCTATTATTGGACAGGAAATGATTAATCTTAGTTGTCTCGAGGACTT 300

Db 241 CATGCAAAATACCTATTATTGGACAGGAAATGATTAATCTTAGTTGTCTCGAGGACTN 300

QY 301 GGAGCCACTGCTGTTGGACTTACTTCAACCATACAGTATGCTCATGGGGTGGAAAT 360

Db 301 GGAGCCACTGCTGTTGGACTTACTTCAACCATACAGTATGCTCATGGGGTGGAAATN 360

QY 361 CAAGGTCAGGCAAGAGAAAAAACAAGTAAAGAGCAATCTCCCACTGACCCGGGAGCAN 420

Db 361 CAAGGTCAGGCAAGAGAAAAAACAAGTAAAGAGCAATCTCCCACTGACCCGGGAGCAN 420

QY 421 AGCACCCTTAGCCCTTACAAAGGACTAGTTCTCTCAAAAATACATGAAACCTTCGGTACC 480

Db 421 AGCACCCTTAGCCCTTACAAAGGACTAGTTCTCTCAAAAATACATGAAACCTTCGGTACC 480

QY 481 CATACTCGGCTGCTGAGCCCTATTAAATACACCCCTCACTCGGCTCCATGAGGCTCAGCC 540

Db 481 CATACTCGGCTGCTGAGCCCTATTAAATACACCCCTCACTCGGCTCCATGAGGCTCAGCN 540

QY 541 CAAAACCTTACTAACTGTTGGATGTCCTCCCTGCACTTTCAGGCCATACATTTCAATC 600

Db 541 CAAAACCTTACTAACTGTTGGATGTCCTCCCTGCACTTTCAGGCCATACATTTCAATN 600

QY 601 CCTGTTCTTGAACAATGGAACAACCTTTCAGCAGAGAAATAAACACCACTTCCCGTTTAGTA 660

Db 601 CCTGTTCTTGAACAATGGAACAACCTTTCAGCAGAGAAATAAACACCACTTCCCGTTTAGTN 660

QY 661 GGACCTCTTGTGTTTCCAATCTGGAATAAACCCATACTCAAACTCCTACCTGTGTAAATTT 720

Db 661 GGACCTCTTGTGTTTCCAATCTGGAATAAACCCATACTCAAACTCCTACCTGTGTAAATTN 720

QY 721 AGCAATACTATAGACACACCAAGCTCCCAATGATCAGTGGGTAAACCTCCACACCA 780

Db 721 AGCAATACTATAGACACACCAAGCTCCCAATGATCAGTGGGTAAACCTCCACACGN 780

QY 781 ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGTACCTCAGCCTATCATTTGTTG 840
Db 781 ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGTACCTCAGCCTATCATTTGTTN 840
QY 841 AATGGCTCTTCAGAACTATATGCTTCTCTCATTTCTTAGTGCCCTATGACCATCTAC 900
Db 841 AATGGCTCTTCAGAACTATATGCTTCTCTCATTTCTTAGTGCCCTATGACCATCTAN 900
QY 901 ACTGAACAGATTTATACATCATGTCTACCTAAGCCCCACAAAGAGATACCAATT 960
Db 901 ACTGAACAGATTTATACATCATGTCTACCTAAGCCCCACAAAGAGATACCAATTN 960
QY 961 CTTCTCTTTCTTATCAGACGAGGTGCTAGGAGACTAGGTACTGGCATTGGCAGTATC 1020
Db 961 CTTCTCTTTCTTATCAGACGAGGTGCTAGGAGACTAGGTACTGGCATTGGCAGTANN 1020
QY 1021 ACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
Db 1021 ACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGGTGACATGGAACNN 1080
QY 1081 GTCACTGACTCCCTGGTGCAGATTCAGATCAACTTAACTCCCTAGCAGCAGTAGTCCITT 1140
Db 1081 GTCACTGACTCCCTGGTGCAGATTCAGATCAACTTAACTCCCTAGCAGCAGTAGTCCNN 1140
QY 1141 CAAAATCGAAGAGCTTTAGACTTGTCTAACCGCAAGAGGGGGAACCTGTTTATTTTAA 1200
Db 1141 CAAAATCGAAGAGCTTTAGACTTGTCTAACCGCAAGAGGGGGAACCTGTTTATTTNN 1200
QY 1201 GGAGAAGACGCTGTTATTTATGTTAATCAATCCAGAAATTCCTAGAGAAAGTTAAAGAA 1260
Db 1201 GGAGAAGACGCTGTTATTTATGTTAATCAATCCAGAAATTCCTAGAGAAAGTTAAAGNN 1260
QY 1261 ATTGAGATCGAATACATCTAGACGAGAGGCTTCAAAACACCGACCTGGGGCTC 1320
Db 1261 ATTGAGATCGAATACATCTAGACGAGAGGCTTCAAAACACCGACCTGGGGCCNN 1320
QY 1321 CTAGGCAATAGGATGCTGGTCTCCCTTCTTAGGAGCTCTAGCAGCTCTAAATATTG 1380
Db 1321 CTAGGCAATAGGATGCTGGTCTCCCTTCTTAGGAGCTCTAGCAGCTCTAAATATNN 1380
QY 1381 TTAATCTCTTTGACCTGTATCTTTAACTCTCTGTTTAAAGTTTGTCTTCCAGAAAT 1440
Db 1381 TTAATCTCTTTGACCTGTATCTTTAACTCTCTGTTTAAAGTTTGTCTTCCAGAAAN 1440
QY 1441 GAAGCTTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
Db 1441 GAAGCTTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481

RESULT 7
RAD41225
ID RAD41225 standard; cDNA; 2074 BP.
XX AC RAD41225;
XX AC
XX AC
DT 30-OCT-2002 (first entry)
XX Human EMBRY-2 cDNA.
XX Human; embryogenesis associated protein; AIDS; reproductive disorder;
XX infertility; endometriosis; endometrial tumour; inflammatory disorder;
XX autoimmune disorder; acquired immune deficiency syndrome; transgenic;
XX ovarian tumour; contact dermatitis; placenta disorder; preclampsia;
XX EMBRY-2; allergy; gene therapy; gene; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
XX CDS 64..1692
FT /*tag= a
FT /product= "EMBRY-2 protein"
FT sig_peptide 64..123

mat_peptide /*tag= b
124..1689 /*tag= c
/product= "Mature EMBRY-2 protein"
WO200248362-A2.
20-JUN-2002.
14-NOV-2001; 2001WO-US043956.
15-NOV-2000; 2000US-0249407P.
(INCY-) INCYTE GENOMICS INC.
Ramkumar J, Arvizu C;
WPI; 2002-537629/57.
P-PSDB; AAE25054.
New polypeptides of human embryogenesis associated proteins for screening modulators useful for treating or preventing disorders e.g. endometriosis, infertility, allergy, preclampsia.
Claim 59; Page 96-97; 97pp; English.
The invention relates to human embryogenesis associated proteins (EMBRY) and nucleic acid molecules encoding such proteins. EMBRY sequences are useful for screening modulators useful for treating or preventing disorders associated with abnormal expression of EMBRY. The disorders treated include reproductive disorders such as infertility, endometriosis, endometrial or ovarian tumour; autoimmune/inflammatory disorder such as acquired immune deficiency syndrome (AIDS), allergies, contact dermatitis; disorders of the placenta such as preclampsia, abruptio placentae etc. Sequences of the invention are also useful for analysing a proteome of a tissue or a cell type. EMBRY proteins are useful as immunogens for preparing antibodies. Polynucleotides of the invention are useful for creating knockin humanised animals or transgenic animals to model human diseases. They are also used in gene therapy. The present sequence is human EMBRY-2 cDNA
Sequence 2074 BP; 583 A; 567 C; 392 G; 531 T; 0 U; 1 Other;
Query Match 93.8%; Score 1389.8; DB 6; Length 2074;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 1424; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
QY 1 ATGGCCCTCCCTTATCATCTTTTCTTCTTACTGTTCTCTTACCCCTTTCGCTCTCACT 60
Db 64 ATGGCCCTCCCTTATTTTCTTCTTACTGTTCTCTTCTCACACCTTTCAGTCTCACT 123
QY 61 GCACCCCTCCCTGCTGTACACACAGTAGTCCCTTACCAAGAGTTTCTATGAAGA 120
Db 124 GCACCCCTCCCTGCTGTACACACAGTAGTCCCTTACCAAGAGTTTCTATGAGA 183
QY 121 ACCGGCTTCTCGAAATATTGATGCCCATCATATAGGAGTTTATTAAGGGAATCC 180
Db 184 ATGGCGCTCCACAGACATATTGATGCCCATCATATAGGAGTTTATTAAGGGAATCC 243
QY 181 ACCTTCACTGCCACACCCATATGCCCACTGTATTAACCTCTGCACTCTTTGTCATG 240
Db 244 GCCTTACCGCCACACCCATATGCCCACTGTATTAACCTCTGCACTCTTTGTCATG 303
QY 241 CATGCAAAATCACTCATTTATTGGACAGGGAATAATGATTAATCCTAGTTGTCTCGAGGACTT 300
Db 304 CATGCAAAATCACTCATTTATTGGACAGGGAATAATGATTAATCCTAGTTGTCTCGAGGACTT 363
QY 301 GGAGCCACTCTGTGTTGGACTTACTTACCCATACCATGATGTTCTGATGGGGTGGAAAT 360
Db 364 GGAGCCACTATCTGTGTTGGACTTACTTACCCATACCATGATGTTCTGATGGGGTGGAGTT 423
QY 361 CAAGGTTCAGGCAAGAGAAACCAAGTAAGGAAGCAATCTCCCAACTGACCCGGGACAT 420

Db 424 CAAGATCAGGCAAGAGAAACATGTAAAGAGATTAATCTCCCACTCACCGGGTACAT 483
Qy 421 AGACACCCCTAGCCCTTACAAAGAGCTAGTGTCTCTCAAAAATCATAGAAACCTCCGTAAC 480
Db 484 AGCACCCCTAGCCCTTACAAAGAGCTAGATCTCTCAAAAATCATAGAAACCTCCGTAAC 543
Qy 481 CATACTCGCTGGTGGCTTATTAATACCAACCTCACTCGGCTCCATGAGGTCTCAGCC 540
Db 544 CATACTCACCTGGTGAAGCCCTATTTAATACCAACCTCACTCGGCTCCATGAGGTCTCGGCC 603
Qy 541 CAAAAACCTACTAAGCTTGGATGTGCTCCCTCCCTGCACCTTCAGGCCATACATTTCAATC 600
Db 604 CAHAACCTACTAAGCTTGGATGTGCTCCCTCCCTGCACCTTCAGGCCATACATTTCAATC 663
Qy 601 CTGTGCTCCTGAAACAATGGAACAACCTTCAGCACGAAATAAACCACTCCGTTTGTAGTA 660
Db 664 CTTGTACTCTGAACAATGGAATAAATTCAGCACGAAATAAACCACTCCATTTTGTAGTA 723
Qy 661 GGACCTCTGTTTCCCAATCTGGAATAAACCACTCAACCTCAACCTCAACCTCAACCTCAACCT 720
Db 724 GGACCTCTGTTTCCCAATCTGGAATAAATCAATCAACCTCAACCTCAACCTCAACCTCAACCT 783
Qy 721 AGCAATACTATAGACACCAACCTCCCAATGATCAGGTGGGTAAACCACTCCCAACCA 780
Db 784 AGCAATACTATAGACACCAACCTCCCAATGATCAGGTGGGTAAACCACTCCCAACCA 843
Qy 781 ATAGTCTGCTACCTCCAGAAATATTTTGTCTGTGTACCTCAGCCTATCAATTTTGTG 840
Db 844 ATAGTCTGCTACCTCCAGAAATATTTTGTCTGTGTACCTCAGCCTATCAATTTTGTG 903
Qy 841 AATGGCTCTTCAGAAATCTAGTCTCTCTCAATCTTAGTGCCCTATGACCATCTAC 900
Db 904 AATGGCTCTTCAGAAATCTAGTCTCTCTCAATCTTAGTGCCCTATGACCATCTAC 963
Qy 901 ACTGAACAAGATTTATACAAATCATGTCTAAGCTTAAAGCCCAACAAAGAGTACCCATT 960
Db 964 ACTGAACAAGATTTATACAAATCATGTCTAAGCTTAAAGCCCAACAAAGAGTACCCATT 1023
Qy 961 CTTCTCTTTGTTATCAGAGCAGAGTGTCTAGGAGAGTGGTACTGGCATTTGGCAGTATC 1020
Db 1024 CTTCTCTTTGTTATCAGAGCAGAGTGTCTAGGAGAGTGGTACTGGCATTTGGCAGTATC 1083
Qy 1021 ACAACCTCTACTCAGTCTACTACAAATCATCTCAAGAAATAAATGGTGAATGGAACAG 1080
Db 1084 ACAACCTCTACTCAGTCTACTACAAATCATCTCAAGAAATAAATGGTGAATGGAACAG 1143
Qy 1081 GTCACCTGACTCCCTGGTGCCTTGAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCCT 1140
Db 1144 GTCACCTGACTCCCTGGTGCCTTGAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCCT 1203
Qy 1141 CAAAATCGAAGAGCTTTAGACTTTGCTAACGCCCAAGAGGGGGAACCTGTGTTTATTTTA 1200
Db 1204 CAAAATCGAAGAGCTTTAGACTTTGCTAACGCCCAAGAGGGGGAACCTGTGTTTATTTTA 1263
Qy 1201 GGAGAAGAACGCTGTTATTTATTTATTTAATCAATCAGAAATTTGCTCAGAGAAATTAAGAA 1260
Db 1264 GGAGAAGAACGCTGTTATTTATTTATTTAATCAATCAGAAATTTGCTCAGAGAAATTAAGAA 1323
Qy 1261 ATTCTGAGATCGAATACAAATCTAGCAGAGAGGAGCTTCAAAACACCGAGCTGGGGCCCTC 1320
Db 1324 ATTCTGAGATCGAATACAAATCTAGCAGAGAGGAGCTTCAAAACACCGAGAGCTGGGGCCCTC 1383
Qy 1321 CTGAGCAATGAGATGCGCTGGTCTCCCTCTCTTAGGAGCTCTAGCAGCTCTAATATTG 1380
Db 1384 CTGAGCAATGAGATGCGCTGGTCTCCCTCTCTTAGGAGCTCTAGCAGCTCTAATATTG 1443
Qy 1381 TTACTCTCTTTGGAGCCCTGTATCTTTTAACCTCTCTTTTAAAGTTTGTCTCTCCAGAAAT 1440
Db 1444 TTACTCTCTTTGGAGCCCTGTATCTTTTAACCTCTCTTTTAAAGTTTGTCTCTCCAGAAAT 1503
Qy 1441 GAAGCTGTAAAGTACAGATGGTCTTACAAATGGAAACCCCA 1481
Db 1504 GAAGCTGTAAAGTACAAATGGTCTTCTCAAAATGGAGCCCA 1544

RESULT 8
AAD24195
ID AAD24195 standard; cDNA; 2930 BP.
XX
AC AAD24195;
XX
XX 07-MAY-2002 (first entry)
DT
XX Human syncytin cDNA.
DE
XX Human; syncytin; preeclampsia; gestational trophoblast disorder;
KW chorioncarinoma; hydatiform mole; placental site tumour; abortion;
KW envelope gene; human endogenous defective retrovirus; HERV-W; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 930..2546
FT /*tag= a
FT /product= "Syncytin"
XX
XX WO200204678-A2.
XX
XX 17-JAN-2002.
XX
XX 09-JUL-2001; 2001WO-US021719.
XX
XX 07-JUL-2000; 2000US-0216657P.
XX
XX (GEMY) GENETICS INST INC.
XX
XX Keith JC, McCoy JM, Mi S;
XX
XX WPI: 2002-171727/22.
XX P-PSDB; AAE14540.
XX
XX Identifying a compound for treating a subject with or at risk of
PT developing preeclampsia, comprises determining whether the expression or
PT activity of syncytin in the cell is modulated in the presence of a test
PT compound.
XX
XX Disclosure; Page 39-42; 43pp; English.
XX
XX The invention relates to identifying compounds which are modulators of
CC syncytin expression. The syncytin modulators are useful in diagnosis and
CC treatment of preeclampsia and gestational trophoblast disorders (e.g.
CC chorioncarinoma, hydatiform mole, placental site tumour and missed/
CC incomplete abortion). Syncytin is a human gene derived from the envelope
CC gene of human endogenous defective retrovirus, HERV-W. The present
CC invention is based partly on the discovery that syncytin expression is
CC dramatically reduced in preeclampsia, and is also mis-localised to the
CC apical syncytiotrophoblast membrane. The present sequence is human
CC syncytin cDNA
XX
SQ Sequence 2930 BP; 842 A; 800 C; 571 G; 717 T; 0 U; 0 Other;

Query Match 90.2%; Score 1335.4; DB 6; Length 2930;
Best Local Similarity 93.9%; Pred. No. 0;
Matches 1390; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

Qy 1 ATGGCCCTCCCTTATCATACTTTCTTTACTTTCTTTACTTTCTTTACCCCTTTTGGTCTCACT 60
Db 930 ATGGCCCTCCCTTATCATATTTTCTTTACTTTCTTTACTTTCTTTTACCCCTTTTCACTCACT 999
Qy 61 GCACCCCTCCATGCTGTATCAACAGTAGTCCCTTACCAAGAGTTTCTATGAAGA 120
Db 990 GCACCCCTCCATGCTGTATGACAGTAGTCCCTTACCAAGAGTTTCTATGAGA 1049
Qy 121 AGCGGCTCTCTGGAAATATTGATGCCCATCATATAGAGTTTATCTAAGGGAATCTCC 180
Db 1050 ATGCAGCGTCCCGGAAATATTGATGCCCATCATGATAGAGTCTTTCTTAAGGGAATCTCC 1109

QY 181 ACCTTCACCTGCCACACCCCATATGCCCCGCAACTGCTATATACTCTGCCACTCTTTTCATG 240
Db |||
QY 1110 ACCTTCACCTGCCACACCCCATATGCCCCGCAACTGCTATATACTCTGCCACTCTTTTCATG 1169
Db |||
QY 241 CATGCAAACTACTCATTTATTTGGACAGGAAATGATTAATCTTAGTTGTCTCTGGAGACTT 300
Db |||
QY 1170 CATGCAAACTACTCATTTATTTGGACAGGAAATGATTAATCTTAGTTGTCTCTGGAGACTT 1229
Db |||
QY 301 GGAGCCACTGTCTGTGGACTTACTTCCACCATACCACTGATGCTCTGATGGGGTGAAT 360
Db |||
QY 1230 GGAGTCACTGTCTGTGGACTTACTTCCACCAACTGGTATGCTGATGGGGTGGAGTT 1289
Db |||
QY 361 CRAAGGTCAAGCAAGAGAAACCAAGTAAGAGAGCAATCTCCCAACTGACCCGGGACAT 420
Db |||
QY 1290 CAAGATCAGCAAGAGAAACCAAGTAAGAGAGTAATCTCCCAACTCACCAGGTACAT 1349
Db |||
QY 421 AGCACCCCTAGCCCTCAAGAGACTAGTTCTCTCAAAACTACATGAAACCCCTCCGTACC 480
Db |||
QY 1350 GGCACCTCTAGCCCTACAAAGACTAGATCTCTCAAACTACATGAAACCCCTCCGTACC 1409
Db |||
QY 481 CATACTGCCCTGTGAGCCCTATTTAATACCACTCACTCGGCTCCATGAGGTCTCAGCC 540
Db |||
QY 1410 CATACTGCCCTGTGAGCCCTATTTAATACCACTCACTCGGCTCCATGAGGTCTCGGC 1469
Db |||
QY 541 CAAAACCTACTAATCTGTGGATGTGCTCCCTGCACTTTCAGGCAATACATTTCAATC 600
Db |||
QY 1470 CAAAACCTACTAATCTGTGGATGTGCTCCCTGCACTTTCAGGCAATATGTTCAATC 1529
Db |||
QY 601 CTGTCTCTGAACAATGGAACAACTTCAGCAGCAAGAAATAAACCACTCCGTTTGTAGTA 660
Db |||
QY 1530 CTGTACTCAACAATGGAACAACTTCAGCAGCAAGAAATAAACCACTCCGTTTGTAGTA 1589
Db |||
QY 661 GGACCTCTGTCTTCCATCTGGAATAACCACTACCTCAACCTCAACCTGTGTAATAATTT 720
Db |||
QY 1590 GGACCTCTGTCTTCCATCTGGAATAACCACTACCTCAACCTCAACCTGTGTAATAATTT 1649
Db |||
QY 721 AGCAATACTATAGACAAACAGCTCCCAATGATCAGGTGGGTAAACCTCCACACGA 780
Db |||
QY 1650 AGCAATACTATAGACAAACAGCTCCCAATGATCAGGTGGGTAAACCTCCACACGA 1709
Db |||
QY 781 ATAGTCTGCCCTACCTCAGGAATATTTTGTCTGTGGTACCTCAGCCTATCATTTGTTG 840
Db |||
QY 1710 ATAGTCTGCCCTACCTCAGGAATATTTTGTCTGTGGTACCTCAGCCTATCATTTGTTG 1769
Db |||
QY 841 AATGGCTCTTCAGAACTATGCTCTCTCTCATCTTCTAGTCCCTATGACCATCTAC 900
Db |||
QY 1770 AATGGCTCTTCAGAACTATGCTCTCTCTCATCTTCTAGTCCCTATGACCATCTAC 1829
Db |||
QY 901 ACTGAACAAGATTTATACAATCATGCTGATCTAAGCCCAACCAAAAGAGTACCCATT 960
Db |||
QY 1830 ACTGAACAAGATTTATACAATCATGCTGATCTAAGCCCAACCAAAAGAGTACCCATT 1889
Db |||
QY 961 CTTCCCTTTGTTATCAGACAGGAGTGTGAGGAGACTAGTACTGCGATTGGCAGTATC 1020
Db |||
QY 1890 CTTCCCTTTGTTATGAGGAGGAGTGTGAGGAGACTAGTACTGCGATTGGCAGTATC 1949
Db |||
QY 1021 ACACCTCTACTCAGTTCTACTACAACTACTCTCAAGAAATAAATGTGACATGGAACAG 1080
Db |||
QY 1950 ACACCTCTACTCAGTTCTACTACAACTACTCTCAAGAAATAAATGTGAGGAGTGAACGG 2009
Db |||
QY 1081 GTCACTGACTCCCTGGTCACTTCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCCT 1140
Db |||
QY 2010 GTCCGCACTCCCTGGTCACTTCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCCT 2069
Db |||
QY 1141 CAAATCGAAGCTTTAGCTTGTCTAACCGCCAAAGAGGGGAACTGTTATTTTAA 1200
Db |||
QY 2070 CAAATCGAAGCTTTAGCTTGTCTAACCGCCAAAGAGGGGAACTGTTATTTTAA 2129
Db |||
QY 1201 GGAGAAGACGCTCTTATGTTAATCAATCCAGAAATGTCTAGTGAAGAGTTAAAGAA 1260
Db |||
QY 2130 GGGGAAGAAATGCTGTTATGTTAATCAATCCGAAATCGTCACTGAGAAAGTTAAAGAA 2189
Db |||

QY 1261 ATTGAGATCGAATACAATGTAGACAGAGAGCTTCAAAACACCAAGCGCTGGGCGCTC 1320
Db |||
QY 2190 ATTGAGATCGAATACAATGTAGACAGAGAGCTTCAAAACACCAAGCGCTGGGCGCTC 2249
Db |||
QY 1321 CTCAGCCATGATGCCCTGGGTCTCCCTCTTTAGGACCTCTAGCAGCTCTAATATTG 1380
Db |||
QY 2250 CTCAGCCATGATGCCCTGGGTCTCCCTCTTTAGGACCTCTAGCAGCTCTAATATTG 2309
Db |||
QY 1381 TTACTCTCTTTGGACCCCTGTATCTTTAACTCTCTTTAGTTTGTCTCTTCCAGAAAT 1440
Db |||
QY 2310 CTACTCTCTTTGGACCCCTGTATCTTTAACTCTCTTTAGTTTGTCTCTTCCAGAAATC 2369
Db |||
QY 1441 GAAGCTGTAAAGCTACAGATGTCTTCAAAATGGAACCCCA 1481
Db |||
QY 2370 GAAGCTGTAAAGCTACAAATGAGCCCAAGATGTCAGTCCAA 2410
Db |||

RESULT 9
AAAX77526

ID AAX77526 standard; cDNA; 2946 BP.

XX AC AAX77526;

XX DT 10-AUG-1999 (first entry)

XX DE Human secreted protein AJ172_2 cDNA.

XX KW Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
XX KW bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
XX KW cell proliferation; cell differentiation; suppressor; tumour inhibition;
XX KW haematopoiesis regulator; activin; inhibin; chemotactic; chemokine;
XX KW haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
XX KW cadherin; tumour invasion suppressor; gene therapy; tissue growth; ss.

XX OS Homo sapiens.

XX XX WO926972-A1.

XX PD 03-JUN-1999.

XX PF 17-NOV-1998; 98WO-US024614.

XX XX 21-NOV-1997; 97US-00976110.

XX PR 18-MAY-1998; 98US-0080478.

XX PR 20-OCT-1998; 98US-00175928.

XX PA (GEMY) GENETICS INST INC.

XX PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;

XX PI Merberg D, Treacy M;

XX PI WPI; 1999-357813/30.

XX PI P-PSDB; AAY08622.

XX XX New polynucleotides encoding secreted proteins.

XX PS Claim 13a; Page 100-101; 142pp; English.

XX CC This invention describes novel human secreted proteins encoded by
XX CC polynucleotides isolated from human adult testes, adult brain, adult
XX CC blood or adult placenta, or murine adult bone marrow or thymus cDNA
XX CC libraries. The products of the invention are predicted to have biological
XX CC activities which would make them suitable for treating, preventing or
XX CC ameliorating medical conditions in humans and animals, although no
XX CC supporting data is given. Suggested activities include nutritional
XX CC activity, cytokine and cell proliferation/differentiation activity,
XX CC immune stimulating (e.g. as vaccines) or suppressing activity,
XX CC haematopoiesis regulating activity, tissue growth activity,
XX CC activin/inhibin activity, chemotactic/chemokine activity, haemostatic
XX CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX CC activity, cadherin/tumour invasion suppressor activity, and tumour
XX CC inhibition activity. The polynucleotides are also stated to be useful for
XX CC gene therapy

XX	Sequence	2946 BP; 858 A; 801 C; 570 G; 717 T; 0 U; 0 Other;	
SY	Query Match	90.2%; Score 1335.4; DB 2; Length 2946;	
	Best Local Similarity	93.9%; Pred. No. 0;	
	Matches 1390; Conservative	0; Mismatches 91; Indels 0; Gaps 0;	
QY	1	ATGGCCCTCCCTTATCATATCTTTCTTTTACTTTCTCTTACCTTCTTACCCCTTTGGCTCTCACT	60
DB	928	ATGGCCCTCCCTTATCATATCTTTCTTTTACTTTCTTTTACCCCTTTTCACTCTCACT	987
QY	61	GCACCCCTCCATGCTGTCTGACAAACAGTAGCTCCCTTACCAAGAGTTTCTATGAAGA	120
DB	988	GCACCCCTCCATGCTGTATGACAGTAGTCTCCCTTACCAAGAGTTTCTATGAGA	1047
QY	121	ACCGGCTTCTCTGGAATATTGATGCCCATCATATAGGAGTTTATCTAAGGGAATCTC	180
DB	1048	ATGCAGCGTCCCGGAATATTGATGCCCATCTGATAGGAGTCTTTCTAAGGGAACCCCC	1107
QY	181	ACCTTCACTGCCACACCCCATATGCCCGCACTGCTATTAACCTCTGCGACTCTTTGCATG	240
DB	1108	ACCTTCACTGCCACACCCCATATGCCCGCACTGCTATCACTCTGCCACTCTTTGCATG	1167
QY	241	CATGCAAACTACTATTATGACAGGGAATAATGATTAATCCTAGTTGTCTCTGGAGACTT	300
DB	1168	CATGCAAACTACTATTATGACAGGGAATAATGATTAATCCTAGTTGTCTCTGGAGACTT	1227
QY	301	GGAGCCACTGCTGTGTTGGACTTACTTCCACCATACCAAGTAGTCTGATGGGGGTGGAATT	360
DB	1228	GGAGTCACTGCTGTGTTGGACTTACTTCCACCCTGATGCTGATGGGGGTGAGTT	1287
QY	361	CAAGGTCAAGCAAGAGAAACAAAGTAAGGAAGCAATCTCCCAACTGACCCGGGACAT	420
DB	1288	CAAGATCAGGCAAGAGAAACATGTAAGGAAGTAATCTCCCAACTCACCCGGGTACAT	1347
QY	421	AGCACCTTAGCCCTACAAAGACTAGTTCTCTCAAACTACATGAAACCTCCGTACC	480
DB	1348	GGCACCTTAGCCCTACAAAGACTAGATCTCTCAAACTACATGAAACCTCCGTACC	1407
QY	481	CATACTCGCTGTGAGCTTATTTAATACCACCTCACTCGGCTCCATGAGGTCTCAGCC	540
DB	1408	CATACTCGCTGTGAGCTTATTTAATACCACCTCACTCGGCTCCATGAGGTCTCGCC	1467
QY	541	CAAAACCTACTAACTGTTGGATGTGCTCCCTCGTCACTTTCAGGCAATCAATTCATC	600
DB	1468	CAAAACCTACTAACTGTTGGATGTGCTCCCTCGTCACTTTCAGGCAATATGTTCAATC	1527
QY	601	CTGTGTTCTGAACAATGGAACAACTTCCAGACAGAAATAACACCACTTCCGTTTGTAGTA	660
DB	1528	CCTGTACCTGAAACAATGGAAACAACCTTCCAGACAGAAATAAACCACTTCCGTTTGTAGTA	1587
QY	661	GGACCTCTTGTTCCTCAATCTGGAATAAACCCATACCTCAAACTCACTGTGTAAATTT	720
DB	1588	GGACCTCTTGTTCCTCAATCTGGAATAAACCCATACCTCAAACTCACTGTGTAAATTT	1647
QY	721	AGCAATACTATAGACAAACCACTCCCAATGATCAGGTGGGTAAACACCTCCCAACAGA	780
DB	1648	AGCAATACTATAGACAAACCACTCCCAATGATCAGGTGGGTAACTCTCCCAACAA	1707
QY	781	ATAGTCTGCCCTACCTCAGGAATATTTTGTCTGTGGTACCTCAGGCTATCATTTGTTG	840
DB	1708	ATAGTCTGCCCTACCTCAGGAATATTTTGTCTGTGGTACCTCAGGCTATCATTTGTTG	1767
QY	841	AATGGCTCTTTCAGAACTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAC	900
DB	1768	AATGGCTCTTTCAGAACTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAC	1827
QY	901	ACTGAACAAGATTATACATATGCTGCTACCTAACGCCCAACCAAGAGTACCCATT	960
DB	1828	ACTGAACAAGATTATACATATGCTGCTACCTAACGCCCAACCAAGAGTACCCATT	1887
QY	961	CTTCCCTTTGTTATCAGACAGAGTCTTAGGAGACTAGGTACTGGCATTGGCAGTATC	1020

DB	1888	CTTCTCTTTTATAGGAGCAGGAGTGTAGGTGCACTAGGTACTGSCATTGGCGTATC	1947
QY	1021	ACAACTCTACTCAGTTCTACTACAAATATCTCAAGAAATAAATGGTGACATGGAACAG	1080
DB	1948	ACAACTCTACTCAGTTCTACTACAAATATCTCAAGAAATAAATGGGACATGGAACGG	2007
QY	1081	GTCACTGACTCCCTGTGTCACTTGGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCTTT	1140
DB	2008	GTCCCGACTCCCTGTGTCACTTGGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCTTT	2067
QY	1141	CRAAATCGAAGAGCTTTAGACTTTGTCTAAACCGCCAAAGAGGGGGAACCTCTTTATTTT	1200
DB	2068	CRAAATCGAAGAGCTTTAGACTTTGTCTAAACCGCTGAAAGAGGGGGAACCTCTTTATTTT	2127
QY	1201	GGAGAAGAACGCTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT	1260
DB	2128	GGAGAAGAACGCTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT	2187
QY	1261	ATTTCGAGATCGAATACAAATGTAGAGCAGAGGAGCTTCAAAACACCGCAACGCTGGGCTC	1320
DB	2188	ATTTCGAGATCGAATACAAATGTAGAGCAGAGGAGCTTCAAAACACCGCAACGCTGGGCTC	2247
QY	1321	CTCAGCAATGATGCTGCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAAATATTG	1380
DB	2248	CTCAGCAATGATGCTGCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAAATATTG	2307
QY	1381	TTACTCTCTTTGGACCTGTATCTTTAACTCTCTTTGTTAAAGTTTGTCTTCCAGAAAT	1440
DB	2308	CTACTCTCTTTGGACCTGTATCTTTAACTCTCTTTGTTAACTCTTCTCTCCAGAAATC	2367
QY	1441	GAAGCTGTAAGCTACAGATGCTTCTTACAAATGGAACCCCA	1481
DB	2368	GAAGCTGTAAGCTACAGATGCTTCTTACAAATGGAACCCCA	2408
RESULT 10			
AAZ59468			
ID	AAZ59468	standard; cDNA; 2946 BP.	
XX	AAZ59468;		
XX	AC		
DT	11-APR-2000	(first entry)	
XX	Human secreted protein Aul72.2 polynucleotide sequence.		
DE	Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;		
KW	placental pathology; metastasis inhibition; nutritional activity;		
KW	immune stimulator; haematopoiesis regulator; tissue growth;		
KW	tumour inhibitor; anti-inflammatory; clone Aul72.2; ATCC_98115;		
KW	gene therapy; ss.		
XX	Homo sapiens.		
XX	OS		
PN	W09960020-A1.		
XX			
PD	25-NOV-1999.		
XX			
PF	17-MAY-1999; 99WO-US010915.		
XX			
PR	18-MAY-1998; 98US-00080478.		
PR	20-OCT-1998; 98US-00175928.		
XX			
PA	(GEM) GENETICS INST INC.		
XX			
PI	Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;		
PI	Herberg D, Mi S, Treacy M;		
XX			
DR	WPI; 2000-116311/10.		
DR	P-PSDB; AAV67313.		
XX			
PT	New polynucleotides encoding secreted cDNA libraries, used to develop		
PT	products for the diagnosis and treatment of neoplastic disease.		
XX			

PS	Claim 14; Page 107-108; 149pp; English.	
XX	This is the human secreted protein AJ172.2 nucleotide sequence, obtained from a human adult testes cDNA library. The invention relates to secreted human and murine proteins. The polynucleotides and proteins are predicted to have biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals. Detection of the levels of the proteins can be used for the diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents which modulate the expression or function of the proteins may be used for treating a neoplastic disease and inhibiting metastasis. Other suggested activities include nutritional activity (e.g. in feeds), cytokine and cell proliferation/differentiation activity, immune stimulating (e.g. as vaccines) or suppressing activity, haematopoiesis regulating activity, tissue growth activity, activin/inhibin activity.	
CC	chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cadherin/tumour invasion suppressor activity, and tumour inhibition activity. The polynucleotide sequences are also stated to be useful for gene therapy	
XX		
SQ	Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 U; 0 Other;	
	Query Match 30.2%; Score 1335.4; DB 3; Length 2946;	
	Best Local Similarity 93.9%; Pred. No. 0;	
	Matches 1390; Conservative 0; Mismatches 91; Indels 0; Gaps 0;	
QY	1 ATGCGCCCTCCCTATACATCTTTCTCTTACTGTTCTCTTACCCCTTTCGCTCTCACT 60	
DB	928 ATGCGCCCTCCCTATACATATTTTCTCTTACTGTTCTTTTACCCCTTTCCTCTCACT 987	
QY	61 GCACCCCTCCCTGCTGTACACAGTAGTCTCCCTTACCAAGATTTCTATGAAGA 120	
DB	988 GCACCCCTCCCTGCTGTATGACCACTAGTCTCCCTTACCAAGATTTCTATGAGA 1047	
QY	121 ACGGGCTTCCTGGAATATTTGATGCCCATATATAGGAGTTTATCTAGGAAATCTCC 180	
DB	1048 ATGACGGCTCCGGAATATTTGATGCCCATATGCTATAGGAGTTTCTTAAAGGAACCC 1107	
QY	181 ACCTTCACTGCCACACCCATATGCGCCGCACTGTATTAACCTCTGCACTCTTTGCAATG 240	
DB	1108 ACCTTCACTGCCACACCCATATGCGCCGCACTGTATTAACCTCTGCACTCTTTGCAATG 1167	
QY	241 CATGCAAAATCTATTAATGGACAGGGAATAATGATTAATCTAGTTGCTCTGGAGGACTT 300	
DB	1168 CATGCAAAATCTATTAATGGACAGGGAATAATGATTAATCTAGTTGCTCTGGAGGACTT 1227	
QY	301 GGACCACTGTCTGTGGACTTACTTCAACCATACAGTATGCTGATGGGGTGGAAAT 360	
DB	1228 GGAGTCACTGTCTGTGGACTTACTTCAACCAACTGGTATGCTGATGGGGTGGAGTT 1287	
QY	361 CAAGTTCAGGCAAGAGAAACAAAGTAAAGGAACAACTCTCCAACTGACCCGGGACAT 420	
DB	1288 CAAGATCAGGCAAGAGAAACAACTGTAAGAGAGTATCTCCAACTCACCCGGGTACAT 1347	
QY	421 AGCACCCCTAGCCCTCAAAAGGACTAGTTCTCTCAAAACTATAGAAACCTCCCGTACC 480	
DB	1348 GGCACCTCTAGCCCTCAAAAGGACTAGTCTCTCAAAACTATAGAAACCTCCCGTACC 1407	
QY	481 CATACTCGCTGGTGAGCTATTTAATACCAACCTCACTCGGCTCCATGAGGTCTCAGCC 540	
DB	1408 CATACTCGCTGGTAAAGCTATTTAATACCAACCTCACTCGGCTCCATGAGGTCTCGGCC 1467	
QY	541 CAAACCCCTACTAATCTGTGGATGTGCTCCCTCCCTGCACTTCAGGCCCATATTTCAATC 600	
DB	1468 CAAACCCCTACTAATCTGTGGATGTGCTCCCTCCCTGCACTTCAGGCCCATATTTCAATC 1527	
QY	601 CCGTGTCTGAAACAAATGGAAACAACTTACGACAGCAAAATAAACAACACTTCGGTTTATGA 660	
DB	1528 CCGTGTCTGAAACAAATGGAAACAACTTACGACAGCAAAATAAACAACACTTCGGTTTATGA 1587	
QY	661 GGACCTCTTGTTCATCTGGAAATACCATACCTCAACCTCACTGCTGTAAATTT 720	
DB	1588 GGACCTCTTGTTCATCTGGAAATACCATACCTCAACCTCACTGCTGTAAATTT 1647	

QY	721 AGCAATATCTATAGACACAAACCAAGCTCCCAATGCATCAGTGGGTAAACACTCCACACGA 780
DB	1648 AGCAATATCTATACATACACAAACCAACTCCCAATGCATCAGTGGGTAACTCTCTCCACACAA 1707
QY	781 ATAGTCTGCTACCTCCCTCAGGAATATTTTGTGTGTGGTACCTCAGCCCTATCATTTCTTTG 840
DB	1708 ATAGTCTGCTACCTCCCTCAGGAATATTTTGTGTGTGGTACCTCAGCCCTATCGTTGTTTG 1767
QY	841 AATGGCTCTTCAGAAATCTATGTGCTTCTCTCTCATTTCTAGTCCCTCATGACCAATCTAC 900
DB	1768 AATGGCTCTTCAGAAATCTATGTGCTTCTCTCTCATTTCTAGTCCCTCATGACCAATCTAC 1827
QY	901 ACTGAACAAGATTTATACAAATCATGTCTGCTACCTAGCCCAACAAACAAAGAGTACCATT 960
DB	1828 ACTGAACAAGATTTATACAAATTTATGTCTATCTTAAGCCCGCAACAAACAAAGAGTACCATT 1887
QY	961 CTTCTCTTTTGTATCAGACGACGAGTGTAGGCAGACTAGGTACTGGCAATTTGGCAGTATC 1020
DB	1888 CTTCTCTTTTGTATAGGACGAGGTGTAGGTGTGCACTAGGTACTGGCAATTTGGCGGTATC 1947
QY	1021 ACAACCTCTACTCAGTTCTTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 1080
DB	1948 ABRACCTCTACTCAGTTCTTACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAG 2007
QY	1081 GTCACCTGACTCCCTGTGTCACTTGTGGAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCCT 1140
DB	2008 GTCGCGCACTCCCTGTGTCACTTGTGGAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCCT 2067
QY	1141 CAAATCGAAGAGCTTTAGACTTGTCTAACCGCCAAAGAGGGGAACCTGTTTATTTTAA 1200
DB	2068 CAAATCGAAGAGCTTTAGACTTGTCTAACCGCTGAAGAGGGGAACCTGTTTATTTTAA 2127
QY	1201 GGAGAAAGACGGCTGTTTATTTATTTATCAATCCAGAAATGTCTACTGAGAAAGTTAAAGAA 1260
DB	2128 GGGAAGAGATGCTGTTTATTTATTTATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2187
QY	1261 ATTGAGATCGAATACAATGTAGACAGAGAGCTTCAAAACACCGAAACGCTGGGGCCTC 1320
DB	2188 ATTGAGATCGAATACAACGTAGACAGAGAGCTTCAAAACACCTGGACCCCTGGGGCCTC 2247
QY	1321 CTCAGCAATGGATGCTCCGGTCTCTCCCTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
DB	2248 CTCAGCAATGGATGCTCCGGATCTCTCCCTCTTAGGACCTCTAGCAGCTCTAATATTG 2307
QY	1381 TTACTCTCTTTGGACCCCTGTATTTAACTCTCTTTTAACTCTCTTTTAACTCTCTCTCCAGAA 1440
DB	2308 CTACTCTCTTTGGACCCCTGTATTTAACTCTCTTTTAACTCTCTCTCTCTCCAGAA 2367
QY	1441 GAAGCTGTAAAGCTACAGATGGTCTTTACAATGGAACCCCA 1481
DB	2368 GAAGCTGTAAAGCTACAAATGGAGCCCAAGATGAGTCCAA 2408

RESULT 11

ADC38776

ID ADC38776 standard; cDNA; 2946 BP.

XX

AC ADC38776;

AC

DT 18-DEC-2003 (first entry)

XX

Human cDNA encoding a secreted protein #63.

DE

XX

ss; gene; immune disorder; severe combined immunodeficiency; SCID;

KW

autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;

KW

rheumatoid arthritis; allergic reaction; asthma; myeloid cell deficiency;

KW

lymphoid cell deficiency; osteoporosis; osteoarthritis;

KW

peripheral nervous system disease; peripheral neuropathy;

KW

Alzheimer's disease; Parkinson's disease; coagulation disorder;

KW

inflammatory disease; systemic inflammatory response syndrome; SIRS;

KW

ischaemia-reperfusion injury; Crohn's disease; anaphylaxis;

KW

hypersensitivity; regeneration; neural cell proliferation; fertility;

KW	tumour; chemokine; human; secreted protein.	QY	1	ATGGCCCTCCCTTATCATACTTTTCTTTTACTGTTCTCTTACCCCTTTCGCTCTCACT	60
XX		DB	928	ATGGCCCTCCCTTATCATATTTTCTTTTACTGTTCTCTTACCCCTTTCGCTCTCACT	987
OS	Homo sapiens.	QY	61	GCACCCCTCCATGCTGCTGACAAACAGTAGTCCCTTACCAAGAGTTTCTATGAAGA	120
XX		DB	988	GCACCCCTCCATGCTGCTGATGACCAAGTAGTCCCTTACCAAGAGTTTCTATGAAGA	1047
XX		QY	121	ACGCGGCTTCTCGAAATATTGATGCCCCATATATAGGAGTTTATTAAGGGAATCC	180
PF	02-APR-2002; 2002US-00114893.	DB	1048	ATGACGCTCCCGAAATATTGATGCCCCATATATAGGAGTTTATTAAGGGAATCC	1107
PR	11-AUG-1995; 95US-005114014.	QY	181	ACCTTCACTGCCACACCCCATATGCCCCGCAACTGCTATTAATCTCTGCACTCTTTG	240
PR	05-APR-1996; 96US-00628364.	DB	1108	ACCTTCACTGCCACACCCCATATGCCCCGCAACTGCTATTAATCTCTGCACTCTTTG	1167
PR	19-APR-1996; 96US-00635311.	QY	241	CATGCAAAATCTATTATTGACAGGGAATATTAATCTCTGCACTCTTTG	300
PR	07-JUN-1996; 96US-00659224.	DB	1168	CATGCAAAATCTATTATTGACAGGGAATATTAATCTCTGCACTCTTTG	1227
PR	17-JUN-1996; 96US-00664596.	QY	301	GGAGCCACTGTCTGTGGACTTACTTCAACCATACCAGTATGTCTGATGGGGTGGAA	360
PR	09-JUL-1996; 96US-00677231.	DB	1228	GGAGTCACTGTCTGTGGACTTACTTCAACCATACCAGTATGTCTGATGGGGTGGAA	1287
PR	26-JUL-1996; 96US-00686878.	QY	361	CAAGTCACTGTCTGTGGACTTACTTCAACCATACCAGTATGTCTGATGGGGTGGAA	420
PR	23-AUG-1996; 96US-00701819.	DB	1288	CAAGTCACTGTCTGTGGACTTACTTCAACCATACCAGTATGTCTGATGGGGTGGAA	1347
PR	27-SEP-1996; 96US-00721488.	QY	421	AGCACCCCTAGCCCTACAAAGACTAGTTCTCTCAAACTACATGAACCCCTCCGTACC	480
PR	27-SEP-1996; 96US-00721798.	DB	1348	GGCACCTCTAGCCCTACAAAGACTAGTTCTCTCAAACTACATGAACCCCTCCGTACC	1407
PR	27-SEP-1996; 96US-00721923.	QY	481	CATACCTCGCTGCTGAGCCCTATTATTAATACACCCCTCACTCGGCTCATAGAGTCT	540
PR	25-OCT-1996; 96US-00738367.	DB	1408	CATACCTCGCTGCTGAGCCCTATTATTAATACACCCCTCACTCGGCTCATAGAGTCT	1467
PR	30-OCT-1996; 96US-00739775.	QY	541	CAAAACCCCTACTAACTGTGGATATGCTCCCTGCACTTCCAGGCTCATAGTTC	600
PR	13-JAN-1997; 97US-00783395.	DB	1468	CAAAACCCCTACTAACTGTGGATATGCTCCCTGCACTTCCAGGCTCATAGTTC	1527
PR	10-APR-1997; 97US-00833823.	QY	601	CTGTGCTTCTGAACTGGAACAACTTTCAGCAGAGAAATTAACACCACTTCCGTTAGTA	660
PR	02-JUN-1997; 97US-00867677.	DB	1528	CTGTGCTTCTGAACTGGAACAACTTTCAGCAGAGAAATTAACACCACTTCCGTTAGTA	1587
PR	05-SEP-1997; 97US-00924838.	QY	661	GGACCTCTTGTGTTCCCAATCTGGAATAACCCATACCTCAAACTCACTGCTGTAATA	720
PR	06-OCT-1999; 99US-00413232.	DB	1588	GGACCTCTTGTGTTCCCAATCTGGAATAACCCATACCTCAAACTCACTGCTGTAATA	1647
XX	(GEMY) GENETICS INST INC.	QY	721	AGCAATCTATAGACAAACCAAGCTCCCAATGCATCAGGTGGGTAAACCTCCACACGA	780
PI	Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;	DB	1648	AGCAATCTATAGACAAACCAAGCTCCCAATGCATCAGGTGGGTAAACCTCCACACGA	1707
PI	Merberg D, Treacy M, Bowman MR, Spaulding V, Carlin-Duckett M;	QY	781	ATAGTCTGCTACCTCCCTCAGGAATATTTTGTCTGTTGTTACCTCAGCCTATCATTT	840
PI	Kelleher K;	DB	1708	ATAGTCTGCTACCTCCCTCAGGAATATTTTGTCTGTTGTTACCTCAGCCTATCATTT	1767
XX	WPI: 2003-657236/62.	QY	841	AATGGCTCTTCAAGATCTATGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	900
DR	P-PSDB; ADC38777.	DB	1768	AATGGCTCTTCAAGATCTATGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1827
XX		QY	901	ACTGAAACAGATTTATACAAATCATGTCGTAACCTTCAAGCCCTCCACCAACCAAGAT	960
PT	Proteins AZ3021 encoded by clone AZ3021 from human adult colon, and	DB	1828	ACTGAAACAGATTTATACAAATCATGTCATATCTTAAGCCCTCCACCAACCAAGAT	1887
PT	BD12716 encoded by clone BD12716 from human fetal kidney cDNA library,	QY	961	CTTCTCTTTTGTATCAGACGAGGAGTCTAGGACAGTACTAGTCTGGCATTGCGATATC	1020
PT	useful for treating e.g. multiple sclerosis and rheumatoid arthritis.	DB	1888	CTTCTCTTTTGTATCAGGACGAGGAGTCTAGGAGTCTAGGAGTCTAGGAGTCTAGG	1947
XX	Disclosure; SEQ ID NO 134; 412pp; English.	QY	1021	ACAACCTCTACTCAGTTCTTACTACAACTATCTCAAGAAATTAATGTCGATGGAACAG	1080
CC	The invention relates to a protein comprising fully defined AZ302 1	DB	1948	ACAACCTCTACTCAGTTCTTACTACAACTATCTCAAGAAATTAATGTCGATGGAACAG	2007
CC	protein or BD127 1 6 protein. The polynucleotides are useful for				
CC	expressing recombinant proteins for analysis and are also useful as				
CC	chromosome markers or tags to identify chromosomes or to map related gene				
CC	positions. The proteins are useful as amino acid supplement, carbon				
CC	source, nitrogen source and carbohydrate source. The proteins are useful				
CC	for treating various immune deficiencies and disorders (e.g. severe				
CC	combined immunodeficiency (SCID)), autoimmune disorders (e.g. multiple				
CC	sclerosis, systemic lupus erythematosus, rheumatoid arthritis), allergic				
CC	reactions (e.g. asthma), myeloid or lymphoid cell deficiencies,				
CC	osteoporosis or osteoarthritis, peripheral nervous system diseases (e.g.				
CC	peripheral neuropathy, Alzheimer's disease, Parkinson's disease),				
CC	coagulation disorders, inflammatory diseases (e.g. Crohn's disease),				
CC	response syndrome (SIRS), ischaemia-reperfusion injury, Crohn's disease),				
CC	anaphylaxis and hypersensitivity. Proteins are also useful for inducing				
CC	tumour immunity, for inducing bone, cartilage, tendon, ligament and/or				
CC	nerve growth or regeneration, for proliferating neural cells and for				
CC	regenerating nerve and brain tissue, for inducing fertility and for				
CC	inhibiting tumour growth. Proteins are also useful as chemokine for				
CC	mammalian cells (e.g. monocytes, fibroblasts, neutrophils), and also				
CC	useful as inhibitors of receptor/ligand interactions. The present				
CC	sequence represents cDNA encoding a human secreted protein.				
XX					
SQ	Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 U; 0 Other;				
	Query Match 90.2%; Score 1335.4; DB 10; Length 2946;				
	Best Local Similarity 93.9%; Pred. No. 0;				
	Matches 1390; Conservative 0; Mismatches 91; Indels 0; Gaps 0;				


```
QY 661 GGACCTCTGTTTCCAAATCTGGAATAACCCATACCTCAAACTCAACCTCACTGTTGTAATTT 720
Db |||
QY 661 GGACCTCTGTTTCCAAATCTGGAATAACCCATACCTCAAACTCAACCTCACTGTTGTAATTT 720
Db |||
QY 721 AGCAATACTATAGACACAACAGCTCCCAATGTCATCAGTGGGTAAACACCTCCACACAGA 780
Db |||
QY 721 AGCAATACTATAGACACAACAGCTCCCAATGTCATCAGTGGGTAACTCTCTCCACACAA 780
Db |||
QY 781 ATAGTCTGCTACCTCAGCAATATTTTGTCTGTTGTTACCTCAGCTATCATTTGTTG 840
Db |||
QY 781 ATAGTCTGCTACCTCAGCAATATTTTGTCTGTTGTTACCTCAGCTATCATTTGTTG 840
Db |||
QY 841 AATGGCTCTTCAGAACTATGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
Db |||
QY 841 AATGGCTCTTCAGAACTATGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
Db |||
QY 901 ACTGAACAAGATTTATACATCTATGTCCTCAAGCCCAACCAAAAGAGTACCCATT 960
Db |||
QY 901 ACTGAACAAGATTTATACATCTATGTCCTCAAGCCCAACCAAAAGAGTACCCATT 960
Db |||
QY 961 CTTCTCTTTTGTATCAGAGCAGAGTGTAGGAGACTAGGTACTGGCATTTGGCAGTATC 1020
Db |||
QY 961 CTTCTCTTTTGTATCAGAGCAGAGTGTAGGAGACTAGGTACTGGCATTTGGCAGTATC 1020
Db |||
QY 1021 ACAACCTCTACTCAGTCTACTACAACTACTCTCAAGAAATAAATGGTGCATGGAACAG 1080
Db |||
QY 1021 ACAACCTCTACTCAGTCTACTACAACTACTCTCAAGAAATAAATGGTGCATGGAACAG 1080
Db |||
QY 1081 GTCACTGACTCCCTGGTGCATCTTGAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCCT 1140
Db |||
QY 1081 GTCCGCGACTCCCTGGTGCATCTTGAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCCT 1140
Db |||
QY 1141 CAAAATCGAAGAGCTTTAGACTTGTATTAATGTTAAATCGGAATCGTCACTGAGAAAGTTAAAGAA 1200
Db |||
QY 1141 CAAAATCGAAGAGCTTTAGACTTGTATTAATGTTAAATCGGAATCGTCACTGAGAAAGTTAAAGAA 1200
Db |||
QY 1201 GGAAGAGACGCTTATTAATGTTAAATCGGAATCGTCACTGAGAAAGTTAAAGAA 1260
Db |||
QY 1201 GGAAGAGACGCTTATTAATGTTAAATCGGAATCGTCACTGAGAAAGTTAAAGAA 1260
Db |||
QY 1261 ATTGAGATCGAATACAACTAGAGCAGAGAGCTTCAAAACACCAACGCTGGGCGCTC 1320
Db |||
QY 1261 ATTGAGATCGAATACAACTAGAGCAGAGAGCTTCAAAACACCAACGCTGGGCGCTC 1320
Db |||
QY 1321 CTGAGCAATGGAATGCTGGTCTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
Db |||
QY 1321 CTGAGCAATGGAATGCTGGTCTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
Db |||
QY 1381 TTAATCTCTTTGAGCCCTGTATCTTAACTCTCTTGAAGTTTGTCTCTCCAGAAAT 1440
Db |||
QY 1381 TTAATCTCTTTGAGCCCTGTATCTTAACTCTCTTGAAGTTTGTCTCTCCAGAAAT 1440
Db |||
QY 1441 GAAGCTGTAAGAGCTACAGATGTCCTTACAAATGGAACCCCA 1481
Db |||
QY 1441 GAAGCTGTAAGAGCTACAAATGGAACCCCAAGATGAGTCCAA 1481
Db |||
```

RESULT 13

ABN97948

ID ABN97948 standard; DNA; 2055 BP.

XX AC

XX ABN97948;

XX XX

XX 01-AUG-2002 (first entry)

XX Human retroviral HERV-7q env coding sequence.

XX DE

XX XX

XX Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;

XX KW

XX multiple sclerosis; ds.

XX OS

XX Human endogenous retrovirus.

XX XX

XX W09967395-A1.

XX FN

XX

PD 29-DEC-1999.

XX

PF 23-JUN-1999; 99WO-FR001513.

XX

PR 23-JUN-1998; 98FR-00007920.

XX

PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX

PI Alliel PM, Perin J, Rieger F;

XX

DR WPI; 2000-160587/14.

XX

New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used for diagnosis, treatment and prevention of autoimmune and neurological diseases.

PT

Claim 3; Page 142-145; 225pp; French.

PS

The present invention relates to new nucleic acid sequences of human endogenous retrovirus, HERV-7q, which is located on chromosome 7q.

CC

Regulatory elements associated with HERV-7q may alter expression of other genes (even remote genes) on the same chromosome, inducing immunological and/or neurological changes (which may be pathological or protective/curative). HERV-7q peptides can be used to improve efficiency of the immune response, e.g. in immunotherapy. HERV-7q peptides and their coding sequences can be used in immunogenic or vaccinating compositions, for protection against autoimmune diseases, particularly multiple sclerosis.

CC

The peptides may also be used (by sequence comparison) to detect/identify endogenous retroviruses that are abnormally expressed in cancer, neuropathologies or other autoimmune diseases. The present sequence was used to illustrate the invention

CC

XX

SQ

Sequence 2055 BP; 576 A; 574 C; 376 G; 529 T; 0 U; 0 Other;

XX

Query Match 90.1%; Score 1333.8; DB 3; Length 2055;

Best Local Similarity 93.8%; Pred. No. 0;

Matches 1389; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY

1 ATGGCCCTCCCTTATCATATCTTTCTTTACTTGTCTCTTACCTCCCTTTTGGCTCTCACT 60

Db

391 ATGGCCCTCCCTTATCATATCTTTCTTTACTTGTCTTTTACCTCTTTTCACTCTCACT 450

QY

61 GCACCCCTCCATGCTGCTGTACAACAGTAGTCCCTTACCAAGATTTCTATGAAGA 120

Db

451 GCACCCCTCCATGCTGCTGTACAACAGTAGTCCCTTACCAAGATTTCTATGAAGA 510

QY

121 ACGGGCTTCTCGGAATATGATGCCCATCATATAGGAGTTTATCTAGGGAACCTCC 180

Db

511 ATGCAGCGTCCCGAAATATGATGCCCATCATATAGGAGTTTATCTAGGGAACCTCC 570

QY

181 ACCTTCACTGCCACACACCATATGCCCGCAACTGTATAAATCTCTGCCACTCTTTGCATG 240

Db

571 ACCTTCACTGCCACACACCATATGCCCGCAACTGTATAAATCTCTGCCACTCTTTGCATG 630

QY

241 CATGCAAAATCTCATTTATTTGACAGGAAAATGATTAATCTCTAGTTGTCTCTGAGGACTT 300

Db

631 CATGCAAAATCTCATTTATTTGACAGGAAAATGATTAATCTCTAGTTGTCTCTGAGGACTT 690

QY

301 GGAGCCACTGTCTGTTGGACTTACTTTCACCCATACAGTATGTCATGGGGTGGAAAT 360

Db

691 GGAGCCACTGTCTGTTGGACTTACTTTCACCCATACAGTATGTCATGGGGTGGAAAT 750

QY

361 CAAGGTCAAGCAAGAGAAAAACAAGTAAAGAGCAATCTCCCAACTGACCCCGGAGCAT 420

Db

751 CAAGATCAGCAGAGAGAAAAACAAGTAAAGAGCAATCTCCCAACTGACCCCGGAGCAT 810

QY

421 AGCACCCTTAGCCCTTACAAAGACTAGTTCTCTCAAAATCATATCAAAACCTTCCTGACC 480

Db

811 GGCACCTCTAGCCCTTACAAAGACTAGATCTCTCAAAATCATATCAAAACCTTCCTGACC 870

QY

481 CATACTCGCCTGTTGAGCTTATTAATACCCCTCACTCGGCTCCTAGGCTCTCAGCC 540

Db 871 CATACTCGCTGGTAAAGCCTATTTAATACACCCCTCACTGGGCTCCTCAGGTCTCGGCC 930
Qy 541 CAAAAACCTACTAATCTGTTGGATGTCCTCCCTGACCTTCAAGGCATACATTTCAATC 600
Db 931 CAAAAACCTACTAATCTGTTGGATGTCCTCCCTGACCTTCAAGGCATATGTTTCAATC 990
Qy 601 CTTGTTCTTGAACAATGGAACAACCTTTCAGCAGACAGAAATAAACACCACTTCCGTTTAAATA 660
Db 991 CTTGTTCTTGAACAATGGAACAACCTTTCAGCAGACAGAAATAAACACCACTTCCGTTTAAATA 1050
Qy 661 GGACCTCTGTTTCCAAATCTGGAATAAACCCATACCTCAAACTCCTGTTGTTAAATTT 720
Db 1051 GGACCTCTGTTTCCAAATCTGGAATAAACCCATACCTCAAACTCCTGTTGTTAAATTT 1110
Qy 721 AGCAATACTATAGACACACACCTCCCAATGATCAGGTGGGTAACTCTCCACACAA 780
Db 1111 AGCAATACTATAGACACACACCTCCCAATGATCAGGTGGGTAACTCTCCACACAA 1170
Qy 781 ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGTACCTCAGCCTATCATTTGTTG 840
Db 1171 ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGTACCTCAGCCTATCATTTGTTG 1230
Qy 841 AATGGCTCTTCAAGATCTATGCTTCTCTCAATCTTAGTGCCCTATGACCATCTAC 900
Db 1231 AATGGCTCTTCAAGATCTATGCTTCTCTCAATCTTAGTGCCCTATGACCATCTAC 1290
Qy 901 ACTGAAACAGATTTATACATATCTGCTACCTAAAGCCCAACAAAGAGTACCCATT 960
Db 1291 ACTGAAACAGATTTATACATATCTGCTACCTAAAGCCCAACAAAGAGTACCCATT 1350
Qy 961 CTTCTCTTTGTTTATCAGAGCAGAGTGTCTAGGAGACTAGGTACTGTCATTTGCGATTC 1020
Db 1351 CTTCTCTTTGTTTATCAGAGCAGAGTGTCTAGGAGACTAGGTACTGTCATTTGCGATTC 1410
Qy 1021 ACAACCTCTACTAGTCTTACTACAACTATCTCAAGAAATAATGTTGATGAGACAG 1080
Db 1411 ACAACCTCTACTAGTCTTACTACAACTATCTCAAGAAATAATGTTGATGAGACAG 1470
Qy 1081 GTCACTGACCTCCCTGGTCACTTCAAGATCACTTAACCTCCCTAGCAGCAGTAGTCCCT 1140
Db 1471 GTCCGCCACTCCCTGGTCACTTCAAGATCACTTAACCTCCCTAGCAGCAGTAGTCCCT 1530
Qy 1141 CAAATCGAAGAGCTTTAGACTTGTCTAAACCCCAAGAGAGGGGAACCTGTTTATTTTA 1200
Db 1531 CAAATCGAAGAGCTTTAGACTTGTCTAAACCCCAAGAGAGGGGAACCTGTTTATTTTA 1590
Qy 1201 GGAGAAACGCTGTTTATTTATGTTAATCAATCCAGATTTGCTCACTGAGAAAGTTAAAGAA 1260
Db 1591 GGGGAAGAATGCTGTTTATTTATGTTAATCAATCCAGATTTGCTCACTGAGAAAGTTAAAGAA 1650
Qy 1261 ATTGAGATCGAATACAAATGTAGAGCAGAGAGCTTCAAAACACCGAACGCTGGGGCCTC 1320
Db 1651 ATTGAGATCGAATACAAATGTAGAGCAGAGAGCTTCAAAACACCGAACGCTGGGGCCTC 1710
Qy 1321 CTGAGCCAAATGATGCTGTTTATTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTT 1380
Db 1711 CTGAGCCAAATGATGCTGTTTATTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTT 1770
Qy 1381 TTACTCTCTTTGACCCCTGATCTTTAACTCCCTGTTTAACTGTTTAACTGTTTAACTGTT 1440
Db 1771 CTACTCTCTTTGACCCCTGATCTTTAACTCCCTGTTTAACTGTTTAACTGTTTAACTGTT 1830
Qy 1441 GAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
Db 1831 GAAGCTGTAAAGCTACAAATGGAACCCCAAGATGAGTCCAA 1871

RESULT 14
ABN97927
ID ABN97927 standard; DNA; 2599 BP.
XX
AC ABN97927;
XX

DT 01-AUG-2002 (first entry)
XX Human retroviral sequence HERV 7 env.
DE Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW multiple sclerosis; ds.
XX Human endogenous retrovirus.
XX WO9967395-A1.
XX 29-DEC-1999.
XX 23-JUN-1999; 99WO-FR001513.
XX 23-JUN-1998; 98FR-00007920.
XX (INRM) INERM INST NAT SANTE & RECH MEDICALE.
XX Alliel PM, Perin J, Rieger F;
XX WPI; 2000-160587/14.
DR
XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
PT for diagnosis, treatment and prevention of autoimmune and neurological
PT diseases.
PS Claim 1; Page 128-129; 225pp; French.
XX The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention
SQ Sequence 2599 BP; 744 A; 718 C; 495 G; 642 T; 0 U; 0 Other;
Query Match 90.1%; Score 1333.8; DB 3; Length 2599;
Best Local Similarity 93.8%; Pred. No. 0;
Matches 1389; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
Qy 1 ATGGCCCTCCCTTATCATACTTTCTCTTACGTCTCTTACCCCTTTCGCTCTCACT 60
Db 929 ATGGCCCTCCCTTATCATATTTTCTTCTTACTGTTCTTTTACCTCTTCTCACTCTCACT 988
Qy 61 GCACCCCTCCATGCTGTACAAACAGTAGTCTCCCTTACCAGAGTTTCTATGAAGA 120
Db 989 GCACCCCTCCATGCTGTATGACCAAGTAGTCTCCCTTACCAGAGTTTCTATGAGA 1048
Qy 121 ACSCGGCTTCTGGAATAATTGATGCCCATATATAGGAGTTTATAGGAAACTCC 180
Db 1049 ATGCAGCGTCCGGAATAATTGATGCCCATATAGGAGTTTCTTATAGGAAACCC 1108
Qy 181 ACCTTCACTGCCACACCCATATGCCCGCACTGTATAACTCTGCACTCTTTCGATG 240
Db 1109 ACCTTCACTGCCACACCCATATGCCCGCACTGTATAACTCTGCACTCTTTCGATG 1168
Qy 241 CATGCAAAATACTCATTTATGACAGGGAATAATTAATCCTAGTTGTCTCGAGGACTT 300
Db 1169 CATGCAAAATACTCATTTATGACAGGGAATAATTAATCCTAGTTGTCTCGAGGACTT 1228
Qy 301 GGAGCCACTGTCTGTTGGACTTACTTCAACCCATACAGTATGCTGATGGGGTGAATT 360
Db 1229 GGAGTCACTGTCTGTTGGACTTACTTCAACCCAACTGGTATGCTGATGGGGTGGAGTT 1288

8059 ACCTTCACTGCGCACACCACATATGCCCCGAACTGCTATCACTCTGCCACTCTTTGCAATG 8118
241 CATGAAATACATCATTTATTTGGACAGGAAATGATTAATCTTAGTGTCTCGAGGACTT 300
8119 CATGCAATACATCATTTATTTGGACAGGAAATGATTAATCTTAGTGTCTCGAGGACTT 8178
301 GGAGCCACTGCTGTTGGACTTACTTCAACCACATACCAGTATGCTGATGGGGTGGAAAT 360
8179 GGAGTCACTGCTGTTGGACTTACTTCAACCACATGGTATGCTGATGGGGTGGAGTT 8238
361 CAAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
8239 CAAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8298
421 AGACCCCTAGCCCTACAAAGACCTAGTCTCTCAAACTACATGAAACCCCTCCGTACC 480
8299 GGACCTCTAGCCCTACAAAGACCTAGTCTCTCAAACTACATGAAACCCCTCCGTACC 8358
481 CATACTCGCTGCTGAGCCTATTATTAACACCCCTCACTCGGCTCCATGAGGTCTCAGCC 540
8359 CATACTCGCTGCTGAGCCTATTATTAACACCCCTCACTCGGCTCCATGAGGTCTCAGCC 8418
541 CAAAACCTACTAATCTGTTGGATGTGCTCCCTGCACTTCAAGGCATACATTTCAATC 600
8419 CAAAACCTACTAATCTGTTGGATGTGCTCCCTGCACTTCAAGGCATATGTTCAATC 8478
601 CTTGTTCTCGAACAATGGAACTTTCAGCAGAGAAATAAACAACCTTCCGTTTGTAGTA 660
8479 CTTGTACTGAACAATGGAACTTTCAGCAGAGAAATAAACAACCTTCCGTTTGTAGTA 8538
661 GGACCTCTCTGTTTCCAACTCGAAATAAACCATACTCAAACTCACTGTGTGTAATTT 720
8539 GGACCTCTCTGTTTCCAACTCGAAATAAACCATACTCAAACTCACTGTGTGTAATTT 8598
721 AGCAATACTATAGACACAAACAGCTCCCAATGCAATGAGTGGGTAACTCTCCACACGA 780
8599 AGCAATACTATAGACACAAACAGCTCCCAATGCAATGAGTGGGTAACTCTCCACACAA 8658
781 ATAGTCTGCTTACCTCAGAGATATTTTGTCTGTTGTTACCTCAGCTATCATTTGTTG 840
8659 ATAGTCTGCTTACCTCAGAGATATTTTGTCTGTTGTTACCTCAGCTATCATTTGTTG 8718
841 AATGGCTCTTCAAGATCTATGTGCTCTCTCTCAATCTTAGTGCCCTCATGACATCTAC 900
8719 AATGGCTCTTCAAGATCTATGTGCTCTCTCTCAATCTTAGTGCCCTCATGACATCTAC 8778
901 ACTGAAACAAGATTATACATCATGTGCTACCTTAAGCCCAACAACAAAGAGTACCCATT 960
8779 ACTGAAACAAGATTATACATCATGTGCTACCTTAAGCCCAACAACAAAGAGTACCCATT 8838
961 CTTCTCTTTTGTATCAGAGCAGGAGTCTAGGCAGACTAGTACTGTCATTGGCAGTATC 1020
8839 CTTCTCTTTTGTATCAGAGCAGGAGTCTAGGCAGACTAGTACTGTCATTGGCAGTATC 8898
1021 ACAACCTCTACTCAGTCTTACTACAACTATCTCAAGAAATAAATGTGTGACATGGAACAG 1080
8899 ACAACCTCTACTCAGTCTTACTACAACTATCTCAAGAAATAAATGTGTGACATGGAACAG 8958
1081 GTCACTGACCTCCCTGGTCACTTCCAGAGTCACTTAACTCCCTAGCAGCAGTGTCTT 1140
8959 GTCCCGGACTCCCTGGTCACTTCCAGAGTCACTTAACTCCCTAGCAGCAGTGTCTT 9018
1141 CAAATCGAAGCTTTAGTACTGCTAACCCCAAGAGAGGGGAACTGTTTATTTTAA 1200
9019 CAAATCGAAGCTTTAGTACTGCTAACCCCTGAAAGAGGGGAACTGTTTATTTTAA 9078
1201 GGAGAAGAACCTGTTATTTATTTAATCAATCCAGAAATGTCACTGAGAAAGTTAAAGAA 1260
9079 GGGAAGAAATCTGTTATTTATTTAATCAATCCAGAAATGTCACTGAGAAAGTTAAAGAA 9138
1261 ATTGAGATCGAATACATGTAGAGCAGAGAGCTTCAAAACACGACGCTGGGGCTC 1320
9139 ATTGAGATCGAATACAAACGTTAGACAGAGAGCTTTCGAAACACTTGGACCTGGGGCTC 9198

QY 1321 CTCAGCCCAATGATGCCCTGGGTTCTCCCTTTCTTAGACCTCTAGACGCTCTAATATTG 1380
Db 9199 CTCAGCCCAATGATGCCCTGGGTTCTCCCTTTCTTAGACCTCTAGACGCTCTAATATTG 9258
QY 1381 TTACTCTCTTTTGGACCCCTGTATCTTTAACTCTTGTAACTCTTGTAGTTGTCTTCCAGAATT 1440
Db 9259 CTACTCTCTTTTGGACCCCTGTATCTTTAACTCTTGTAACTCTTGTAGTTGTCTTCCAGAATC 9318
QY 1441 GAAGCTGTAAAGCTACAGATGGTCTTTACAAATGGAACCCCA 1481
Db 9319 GAAGCTGTAAACTAAATGAGGCCCAAGATGCGAGTCCAA 9359
RESULT 16
ABL61744
ID ABL61744 standard; DNA; 56093 BP.
XX ABL61744;
AC ABL61744;
XX
DT 15-MAY-2002 (first entry)
XX
DE Colon adenocarcinoma related gene sequence SEQ ID NO:81.
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX Homo sapiens.
XX
PN W0200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US010838.
XX
PR 05-JUN-2000; 2000US-0209473P.
PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0233133P.
PR 18-SEP-2000; 2000US-0233617P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237294P.
PR 02-OCT-2000; 2000US-0237295P.
PR 02-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.

PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
PI
XX WPI; 2002-198264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
PT
XX Claim 1; SEQ ID NO 81; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 847 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms's
CC tumour
XX
XX Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16881 T; 0 U; 0 Other;
SQ
Query Match 90.1%; Score 1333.8; DB 6; Length 56093;
Best Local Similarity 93.8%; Pred. No. 0;
Matches 1389; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
QY 1 ATGGCCCTCCCTTATCATATTTTCTCTTACTGTTCTTTACCCCTTTTCCTCTCACT 35938
DB 35879 ATGGCCCTCCCTTATCATATTTTCTCTTACTGTTCTTTACCCCTTTTCCTCTCACT 35938
QY 61 GCACCCCTCCATGCTGCTGTACAAACAGTAGTCTCCCTTACCAAGAGTTTCTATGAAGA 120
DB 35939 GCACCCCTCCATGCTGCTGTATGACCAAGTAGTCTCCCTTACCAAGAGTTTCTATGAGA 35998
QY 121 ACAGGCTCTCGAATATTTGATGCTCCCATCATATAGGAGTTTATCTAGGGAATCTCC 180
DB 35999 ATGACGGTCCCGAATATTTGATGCTCCCATCATATAGGAGTTTCTTAAGGGAATCTCC 36058
QY 181 ACCTTCACTCCCAACCCATATGCTCCCGCAACTGCTATTAATCTGCACTCTTTGCGATG 240
DB 36059 ACCTTCACTCCCAACCCATATGCTCCCGCAACTGCTATTAATCTGCACTCTTTGCGATG 36118
QY 241 CATGCAAAATCTCATATTGAGCAGGGAAAATGATTAATCTAGTTGCTCTGAGGACTT 300
DB 36119 CATGCAAAATCTCATATTGAGCAGGGAAAATGATTAATCTAGTTGCTCTGAGGACTT 36178
QY 301 GGAGCCACTCTGTTGGACTTACTTCAACCATACCATGATGCTGATGGGGTGGAAAT 360
DB 36179 GGAGTCACTGTTGTTGGACTTACTTCAACCAAACTGGTATGCTGATGGGGTGGAGTT 36238
QY 361 CAAGGTGAGGCAAGAGAAAAAAGTAAGAGAGCAATCTCCCAACTGACCCCGGGGACAT 420
DB 36239 CAAGATCAGGCAAGAGAAAAAAGTAAGAGAGCAATCTCCCAACTGACCCCGGGTACAT 36298
QY 421 AGCACCCCTAGCCCTTCAAAAGGACTAGTTTCTCTCAAAACTATCATGAACCCCTCCGTAAC 480

AAF55630
ID AAF55630 standard; DNA; 2781 BP.
XX AC
XX AC
XX AAF55630;
XX DT
XX 29-MAY-2001 (first entry)
XX Nucleotide sequence of a human endogenous retrovirus envelope protein.
DE DE
XX Envelope protein; HERV; syncytia formation; placental development;
KW syncytia; cancer; cell adhesion; ss.
XX OS
XX Human endogenous retrovirus.
XX Key Location/Qualifiers
FT CDS 762..2378
FT /*tag= a
FT /product= "envelope protein"
XX WO200116171-A1.
XX
XX 08-MAR-2001.
XX 01-SEP-2000; 2000WO-FR002429.
XX 01-SEP-1999; 99FR-00011141.
XX 15-SEP-1999; 99FR-00011793.
XX (INMR) BIO MERIEUX.
XX (INMR) INST NAT SANTE & RECH MEDICALE.
XX Mallet F, Cosset F, Blond J, Lavillette D, Bouton O, Ruggieri A;
XX WPI; 2001-226676/23.
XX P-PSDB; AAB67652.
XX Detecting expression of human endogenous retrovirus envelope protein in
XX cells of a tissue or culture, from its ability to induce syncytia.
XX
XX Disclosure; Page 44-45; 57pp; French.
XX
XX The present sequence encodes a human endogenous retrovirus envelope
XX protein. The specification describes a method for detecting expression of
XX an envelope protein from a human endogenous retrovirus (HERV), in cells,
XX of a tissue or culture. The method comprises detecting syncytia formation
XX due to the fusogenic properties of the envelope protein. Envelope
XX polypeptides and polynucleotides are used to produce therapeutic or
XX prophylactic compositions, particularly for treatment of cancer, to
XX correct defects in placental development (or other natural formation of
XX other types of syncytia), and to promote adhesion of cells in grafts or
XX cellular repair processes. Expression of sequences antisense to the
XX polynucleotide are used to prevent formation of syncytia
XX
XX Sequence 2781 BP; 747 A; 770 C; 556 G; 708 T; 0 U; 0 Other;
XX
XX Query Match 89.8%; Score 1330.6; DB 5; Length 2781;
XX Best Local Similarity 93.7%; Pred. No. 0;
XX Matches 1387; Conservative 94; Mismatches 94; Indels 0; Gaps 0;
XX
XX 1 ATGGCCCTCCCTTATCATACCTTTCTCTTACTGTTCTCTTACCCCTTTTCCTCTCACT 60
XX 762 ATGGCCCTCCCTTATCATATTTTCTCTTACTGTTCTTTTACCCTCTTTTCACTCTCACT 821
XX
XX 61 GCACCCCTCCCTGCTGTGTACACAGTAGTCTCCCTTACCAGAGTTTCTATGAAGA 120
XX 822 GCACCCCTCCCTGCTGTGTATGACAGTAGTCTCCCTTACCAGAGTTTCTATGAAGA 881
XX
XX 121 AGCGGGCTTCTGGAAATATTCATGCCCATCATATAGAGTTTATCAAGGAATCTCC 180
XX 882 ATGCAGGCTCCCGAAATATTCATGCCCATCGTATAGAGTCTTCTTAAGGGAAACCC 941
XX 181 ACCTTCAGTCCCAACCCATATGCGGCACTGCTATTAATCTTGGCACTCTTTGCAATG 240
XX
Db 942 ACCTTCAGTCCCAACCCATATGCGGCACTGCTATCACTCTGCCACTCTTTGCAATG 1001
Qy 241 CATGCAAACTACTCAATTTATGACAGGGAATAATGATTAATCCTAGTTGCTGGAGACTT 300
Db 1002 CATGCAAACTACTCAATTTATGACAGGGAATAATGATTAATCCTAGTTGCTGGAGACTT 1061
Qy 301 GGAGCCACTGCTGTTGGACTTACTTCAACCATACCAGTATGCTCTGATGGGGTGAAT 360
Db 1062 GGAGTCACTGCTGTTGGACTTACTTCAACCAACTGGTATGCTCTGATGGGGTGAAT 1121
Qy 361 CAAGGTCAGGCAAGAGAAACAACTAAAGGAACAATCTCCCAACTGACCCGGGGAAT 420
Db 1122 CAAGATCAGGCAAGAGAAACAACTAAAGGAACAATCTCCCAACTGACCCGGGGAAT 1181
Qy 421 AGCACCCCTAGCCCTTACAAAGGACTAGTTCTCTCAAACTACATGAACCCCTCGTACC 480
Db 1182 GGCACCTCTAGCCCTTACAAAGGACTAGATCTCTCAAACTACATGAACCCCTCGTACC 1241
Qy 481 CATACTCGCTGGTGAGCCTATTTAATACCACTCTCACTCGGCTCCATGAGTCTCAGCC 540
Db 1242 CATACTCGCTGGTGAAGCCTATTTAATACCACTCTCACTCGGCTCCATGAGTCTCGGCC 1301
Qy 541 CAAAACCCCTACTAACTGTTGGATGTCCTCCCTCGCTCTCAGGCCATACATTTCAATC 600
Db 1302 CAAAACCCCTACTAACTGTTGGATGTCCTCCCTCGCTCTCAGGCCATATGTTCAATC 1361
Qy 601 CCTGTTCTGAAACAATGGAACAACCTTCAGACAGAAATAAACACCACTCTCCGTTTACTA 660
Db 1362 CCTGTACTGAAACAATGGAACAACCTTCAGACAGAAATAAACACCACTCTCCGTTTACTA 1421
Qy 661 GGACCTCTGTTTCCAACTCTGGAATAACCCATACCTCAAACTCACTGCTGTAATAATTT 720
Db 1422 GGACCTCTGTTTCCAACTCTGGAATAACCCATACCTCAAACTCACTGCTGTAATAATTT 1481
Qy 721 AGCAATACTATAGACAAACCCAGCTCCCAATGCATCAGTGGGTAAACCTCCCAACA 780
Db 1482 AGCAATACTATAGACAAACCCAGCTCCCAATGCATCAGTGGGTAAACCTCCCAACA 1541
Qy 781 ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGTGTACCTCAGCTCTCATTTGTTG 840
Db 1542 ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGTGTACCTCAGCTCTCATTTGTTG 1601
Qy 841 AATGGCTCTTCAAGATCTATGCTCTCTCTCATTTCTAGTGGCCCTATGACCACTAC 900
Db 1602 AATGGCTCTTCAAGATCTATGCTCTCTCTCATTTCTAGTGGCCCTATGACCACTAC 1661
Qy 901 ACTGAACAAGATTTATACAATCATGTCGTAACCTTAAGCCCAACAACCAAGAGTACCATT 960
Db 1662 ACTGAACAAGATTTATACAGTTTATGTATATCTTAAGCCCAACAACCAAGAGTACCATT 1721
Qy 961 CTTCTCTTTGTTATCAGCAGGAGTGTAGGAGAGTACTAGTACTGGCATTTGGCAGTATC 1020
Db 1722 CTTCTCTTTGTTATAGGAGGAGTGTAGGAGAGTACTAGTACTGGCATTTGGCAGTATC 1781
Qy 1021 ACAACCTCTACTCAGTTTCTACTACAACTATCTCAAGAAATAAATGGTGACATGGAAACAG 1080
Db 1782 ACAACCTCTACTCAGTTTCTACTACAACTATCTCAAGAAATAAATGGGACATGGAAACAG 1841
Qy 1081 GTCACTGACTCCCTGGTCACTTTGCAAGATCACTTAACCTCTCCCTAGCAGCAGTAGTCTT 1140
Db 1842 GTGCGGCACTCCCTGGTCACTTTGCAAGATCACTTAACCTCTCCCTAGCAGCAGTAGTCTT 1901
Qy 1141 CAAAATCGAAGAGCTTTTAGACTTTGTAACCGCCAAAGAGGGGGAACCTGTTATTTTAA 1200
Db 1902 CAAAATCGAAGAGCTTTTAGACTTTGTAACCGCCGAAAGAGGGGGAACCTGTTATTTTAA 1961
Qy 1201 GGAGAAGAAACGCTGTTATTTATGTTAATCAATCCAGAAATGTCATGAGAAAGTTAAAGAA 1260
Db 1962 GGGGAAGAAATGCTGTTATTTATGTTAATCAATCCGAAATCGTCACTGAGAAAGTTAAAGAA 2021
Qy 1261 ATTTCAGATCGAATACAAATGTAGACGAGAGGCTTCAAAACACCGAAGCTGGGGCTC 1320
Db 2022 ATTTCAGATCGAATACAAACGTTAGACGAGAGGCTTCGAAACACTTGGACCTGGGGCTC 2081

QY 1321 CTCAGCCATGGATGCCCTGGTTCTCCCTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
 DB 2082 CTCAGCCATGGATGCCCTGGATTCTCCCTCTTAGGACCTCTAGCAGCTATAATATTG 2141
 QY 1381 TTACTCTCTTTGACCTGTATCTTTAACTCTCTTTGAAGTTGTCTTCCAGAATT 1440
 DB 2142 CTACTCTCTTTGACCTGTATCTTTAACTCTCTTTGAAGTTGTCTTCCAGAAATC 2201
 QY 1441 GAAGCTGTTAAAGTACAGATGGTCTTACAAATGGAACCCCA 1481
 DB 2202 GAAGCTGTTAAAGTACAAATGGAGCCCAAGATGCAATGTCCTCA 2242

RESULT 18

AAA59215
 ID AAA59215 standard; DNA; 7582 BP.

AC AAA59215;

DT 07-NOV-2000 (first entry)

XX Human endogenous retrovirus W (HERV-W) sequence.

DE Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;

KW gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.

XX Human endogenous retrovirus.

PH Key Location/Qualifiers

FT LTR 1. .120

FT /*tag= a

FT /note= "R of 5' LTR"

FT LTR 121. .575

FT /*tag= b

FT /note= "US of 5' LTR"

FT primer_bind 579. .596

FT /*tag= c

FT CDS 5581. .7194

FT /*tag= d

FT /note= "ORF1 env538"

FT CDS 7039. .7194

FT /*tag= e

FT /note= "ORF2 52 AA"

FT CDS 7112. .7255

FT /*tag= f

FT /note= "ORF3 48 AA"

FT misc_feature 7244. .7254

FT /*tag= g

FT /note= "polypurine tract"

FT LTR 7256. .7582

FT /*tag= h

FT /note= "U3-R of 3' LTR"

FT polyA_signal 7563. .7569

FT /*tag= i

XX WO200043521-A2.

PN 21-JAN-2000; 2000WO-FR000144.

PD 27-JUL-2000.

XX 21-JAN-2000; 2000WO-FR000144.

XX 21-JAN-1999; 99FR-00000888.

XX (INMR) BIO MERIEUX.

XX Paranhos-Baccala G, Mallet F, Voisset C;

XX WPI; 2000-499229/44.

XX New nucleic acid from human endogenous retrovirus, useful e.g. for

XX diagnosis of autoimmune disease and complications of pregnancy, contains

XX at least part of the gag gene.

XX

XX

XX

XX

XX

XX

XX Disclosure; Page 49-52; 53pp; French.

XX The present sequence represents an endogenous retrovirus, which is
 CC associated with an autoimmune disease, and is integrated into the human
 CC genome. The retrovirus is human endogenous retrovirus W (HERV-W). The
 CC HERV-W retrovirus is associated with autoimmune disease, failure of
 CC pregnancy or disorders of pregnancy. HERV-W nucleic acid fragments, or
 CC proteins derived from it, are useful for diagnosis of autoimmune disease
 CC (specifically multiple sclerosis) and for monitoring pregnancy. The
 CC nucleic acid fragments may also be used for in situ labelling of isolated
 CC chromosomes, while the transcription product can be used to study or
 CC monitor T cell proliferation in vitro

XX SQ Sequence 7582 BP; 2156 A; 1876 C; 1538 G; 1796 T; 0 U; 216 Other;

Query Match 89.4%; Score 1324.2; DB 3; Length 7582;

Best Local Similarity 92.2%; Pred. No. 0;

Matches 1365; Conservative 24; Mismatches 92; Indels 0; Gaps 0;

QY 1 ATGGCCCTCCCTTATCATCTTTCTTTACTGTTCTTTACCTCTTACCTCTTCTGCTCACT 60

DB 5581 ATGGSCCTCCCTTATCATATTTTCTCTKTASTGTTSTTTTACCTCTTCTCACTCACT 5640

QY 61 GCACCCCTCCATGCTGTACACACAGTAGTCCCTTACCAAGAGTTCTTATGAGA 120

DB 5641 GCACCCCTCCATGCTGTATGACACAGTAGTCCCTTACCAAGAGTTCTTATGAGA 5700

QY 121 AGCGGCTTCTTGGAAATATTGATGCCCATATATAGGAGTTTATCTAAGGGAATCC 180

DB 5701 ATGACGCTCCCGAAATATTGATGCCCATATGATAGGAGTTTATCTAAGGGAATCC 5760

QY 181 ACCTTCACTGCCACACCCCATATGCCCGCACTGTATTAACCTCTGCCACTCTTTGCATG 240

DB 5761 ACCTTCACTGCCACACCCCATATGCCCGCACTGTATTAACCTCTGCCACTCTTTGCATG 5820

QY 241 CATGCAATACTCATTTATTGGACAGGAAATGATTAATCTTCTGAGGACTT 300

DB 5821 CATGCAATACTCATTTATTGGACAGGAAATGATTAATCTTCTGAGGACTT 5880

QY 301 GGAGCCACTGTCTGTGGACTTACTTACCCATACCACTATGCTCATGGGGTGAAT 360

DB 5881 GGAGTCACTGTCTGTGGACTTACTTACCCATACCACTATGCTCATGGGGTGAAT 5940

QY 361 CAAGGTCAAGCAAGAGAAAAACAAGTAAAGGAGCAATCTCCCAACTGACCCGGGACAT 420

DB 5941 CAAGATCAGCAAGAGAAAAACAAGTAAAGGAGTAACTCTCCCAACTGACCCGGGACAT 6000

QY 421 AGCACCTTAGCCCTTACAAAGACTAGTCTCTCAAACTACATGAACCTTCCGTACC 480

DB 6001 GGCACCTTAGCCCTTACAAAGACTAGTCTCTCAAACTACATGAACCTTCCGTACC 6060

QY 481 CATACTCGCTGTGAGCCTATTATACCCCTCACTCGGCTCATGAGGTCTCAGCC 540

DB 6061 CATACTCGCTGTGAGCCTATTATACCCCTCACTCGGCTCATGAGGTCTCAGCC 6120

QY 541 CAAAACCTTAACTGTTGGATGTCCTCCCTGCACCTTCAGGCAATATTTCAATC 600

DB 6121 CAAAACCTTAACTGTTGGATGTCCTCCCTGCACCTTCAGGCAATATTTCAATC 6180

QY 601 CTTGTTCTTGAAATGGAACAACTTTCAGACAGAAATAAACCACTTCCGTTTGTAGTA 660

DB 6181 CTTGTTCTTGAAATGGAACAACTTTCAGACAGAAATAAACCACTTCCGTTTGTAGTA 6240

QY 661 GGACCTCTTGTTCCTCAATCTGGAAATACCCATACCTCAAACTCACTGTGTAATTTT 720

DB 6241 GGACCTCTTGTTCCTCAATCTGGAAATACCCATACCTCAAACTCACTGTGTAATTTT 6300

QY 721 AGCAATACTATAGACACAACTGCTCCCAATGATCAGGTGGGTAACACCTCCACACGA 780

DB 6301 AGCAATACTATAGACACAACTGCTCCCAATGATCAGGTGGGTAACACCTCCACACGA 6360

QY 781 ATAGTCTGCCTACCTCAGGAAATATTTTGTCTGTGGTACCTCAGCCTATCATTTGTTG 840

Db 6361 ATAGTCTGCTACCTCAGGAATATTTTGTGTGGTACCTCAGCTATCGTTGTTG 6420
QY 841 AATGGCTCTTCAAGTCTATAGTCTTCTCTCTATCTTCTAGTGGCCCTATGACCACTAC 900
Db 6421 AATGGCTCTTCAAGTCTATAGTCTTCTCTCTATCTTCTAGTGGCCCTATGACCACTAC 6480
QY 901 ACTGAACAGATTATACATCATGTCGTACCTTAAGCCCCACAAACAAAGAGTACCCATT 960
Db 6481 ACTGAACAGATTATACATCATGTCGTACCTTAAGCCCCACAAACAAAGAGTACCCATT 6540
QY 961 CTTCTCTTTTGTATACAGACAGAGTGTAGGACAGACTAGGTACTGGCAATTGGCAGTATC 1020
Db 6541 CTTCTCTTTTGTATAGGACAGAGTGTAGGACAGACTAGGTACTGGCAATTGGCAGTATC 6600
QY 1021 ACAACCTCTACTCAGTCTTACTACAACTATCTCAAGAAATAAATGGTGACATGGACAG 1080
Db 6601 ACAACCTCTACTCAGTCTTACTACAACTATCTCAAGAAATAAATGGTGACATGGACAG 6660
QY 1081 GTCACTGACCTCCCTGGTCACTTGCAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
Db 6661 GTGCGCCGACTCCCTGGTCACTTGCAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT 6720
QY 1141 CAAAATCGAAGAGCTTTAGACTTGTAAACCGCCAAAGAGGGGGAACCTGTTTATTTTA 1200
Db 6721 CAAAATCGAAGAGCTTTAGACTTGTAAACCGCTGAGAGGGGGAACCTGTTTATTTTA 6780
QY 1201 GGAAGAAACGCTGTATTATGTTAATCAATCCAGAAATGTGCTACTGAGAAAGTTAAGAA 1260
Db 6781 GGAAGAAACGCTGTATTATGTTAATCAATCCAGAAATGTGCTACTGAGAAAGTTAAGAA 6840
QY 1261 ATTGAGATCGAATACAAATGTAGAGCAGAGAGCTTCAAAACACCGAACCTGGGGCTC 1320
Db 6841 ATTGAGATCGAATACAAATGTAGAGCAGAGAGCTTCAAAACACCGAACCTGGGGCTC 6900
QY 1321 CTAGCCCAATGGATGCGCTGGTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
Db 6901 CTAGCCCAATGGATGCGCTGGTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 6960
QY 1381 TTACTCTCTTTGGACCTGTATCTTAACTCTCTTAACTTGTCTTCTCAGAAAT 1440
Db 6961 CTACTCTCTTTGGACCTGTATCTTAACTCTCTTAACTTGTCTTCTCAGAAAT 7020
QY 1441 GAAGCTGTAAAGTACAGATGGTCTTACAAATGGAACCCCA 1481
Db 7021 GAAGCTGTAAAGTACAGATGGTCTTACAAATGGAACCCCA 7061

RESULT 19
ID AAX25665 standard; cDNA to mRNA; 7582 BP.
XX AC AAX25665;
XX AC AAX25665;
XX 21-MAY-1999 (first entry)
XX Complete human endogenous retrovirus W genome.
XX Clone; human endogenous retrovirus; genome; autoimmune disease;
XX multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
XX disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
XX Human endogenous retrovirus.
XX OS
XX WO9902696-A1.
XX 21-JAN-1999.
XX 06-JUL-1998; 98WO-PR001442.
XX 07-JUL-1997; 97FR-00008815.
XX (INMR) BIO MERIEUX.

XX Beseme F, Blond J, Bouton O, Mandrand B, Mallet F;
XX WPI; 1999-120897/10.
XX New nucleic acid sequences from human endogenous retrovirus-W - expressed
XX exclusively in placenta and useful in diagnosis and therapy of autoimmune
XX disease, and abnormal or failed pregnancy.
XX Claim 1; Page 71-74; 106pp; French.
XX This sequence represents the complete sequence of the human endogenous
XX retrovirus (HERV) W genome. The nucleic acids, their fragments or
XX peptides encoded by them are markers of autoimmune disease (e.g. multiple
XX sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,
XX insulin-dependent diabetes and related pathologies) and of abnormal or
XX unsuccessful pregnancy and can be used as chromosomal markers for
XX susceptibility to these conditions, or proximity markers of genes
XX associated with this susceptibility
XX Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 Other;
SQ

Query Match 89.3%; Score 1322.6; DB 2; Length 7582;
Best Local Similarity 92.1%; Pred. No. 0;
Matches 1364; Conservative 24; Mismatches 93; Indels 0; Gaps 0;
QY 1 ATGGCCCTCCCTTATCATCTTTCTTCTTACTGTCTTCTTACCCCTTTCCTCTCACT 60
Db 5581 ATGGCCCTCCCTTATCATATTTTCTTCTTASTGTSTTTTACCTSTTTCCTCTCACT 5640
QY 61 GCACCCCTCCATGCTGTGTACACACAGTAGTCTCCCTTACCAAGAGTTTCTATGAAGA 120
Db 5641 GCACCCCTCCATGCTGTGTATGACAGTAGTCTCCCTTACCAAGAGTTTCTATGAAGA 5700
QY 121 AGCGGCTTCTTGGAAATATTGATGCCCATCATATAGAGTTTATCTAAGGAACTCC 180
Db 5701 ATGCAGGCTCCCGAAATATTGATGCCCATCGTATAGAGTCTTSTTAAGGAACTCC 5760
QY 181 ACCTTCACTGCCCCACACCCATATGCCGCACTGCTATAACTCTGCCACTCTTTGCAATG 240
Db 5761 ACCTTCACTGCCCCACACCCATATGCCGCACTGCTATACTCTGCCACTCTTTGCAATG 5820
QY 241 CATGCAAACTACTCATTTATTTGGACAGGAAATGATTAATCTTAGTTGTCTCGAGGACTT 300
Db 5821 CATGCAAACTACTCATTTATTTGGACAGGAAATGATTAATCTTAGTTGTCTCGAGGACTT 5880
QY 301 GGAGCCACTGTCTGTGGACTTATTTCAACCCATACAGTATGTCTCATGGGGTGAATTT 360
Db 5881 GGAGTCACTGTCTGTGGACTTATTTCAACCCATACAGTATGTCTCATGGGGTGAATTT 5940
QY 361 CAAGTTCAGGACAGAGAAACAAAGTAAAGGAGCAATCTCCCACTGACCCCGGACAT 420
Db 5941 CAAGATCAGGACAGAGAAACAAAGTAAAGGAGCAATCTCCCACTGACCCCGGACAT 6000
QY 421 AGCACCCCTAGCCCTACAAAGGACTAGTTCTCTCAAACTACATGAAACCTCTCGTACC 480
Db 6001 GGACACCTTAGCCCTACAAAGGACTAGATCTCTCAAACTACATGAAACCTCTCGTACC 6060
QY 481 CATACTCGCTGTGTAGCCCTATTTATACACCTCATCTGGCTCCATGAGGCTCAGCC 540
Db 6061 CATACTCGCTGTGTAGCCCTATTTATACACCTCATCTGGCTCCATGAGGCTCAGCC 6120
QY 541 CAAAACCTACTAATCTGTGTGATGTCTCCCTGCACTTTCAGGCAATACATTTCAATC 600
Db 6121 CAAAACCTACTAATCTGTGTGATGTCTCCCTGCACTTTCAGGCAATACATTTCAATC 6180
QY 601 CTTGTTCTGACAAATGGAACAACTTTCAGCAAGAAATAAACACCTCCGTTTGTAGTA 660
Db 6181 CTTGTTCTGACAAATGGAACAACTTTCAGCAAGAAATAAACACCTCCGTTTGTAGTA 6240
QY 661 GGACCTCTTGTTCCTCAATCTGGAATAACCATACCTCAACCTCAGCTGTGTAAATTT 720
Db 6241 GGACCTCTTGTTCCTCAATCTGGAATAACCATACCTCAACCTCAGCTGTGTAAATTT 6300

D _b	763	 ATGCGCCTCCCTTATCATATTTTCTCTAGTGTTCTTTCAACCCTGTGTTTCACCTCTCACT	822
Q _y	61	GCAACCCCTCCATGCTGCCTGTGTACACCACTAGCTCCCCCTTACCAAGAGTTTCTATGAAGA	120
D _b	823	GCACCCCTCCATGCCGCTGTATGACACGATAGTCCCCTCACCCAGAGTTTCTATGGAGA	882
Q _y	121	ACGGGCTTCTGGAAATATTGATGCCCCCATCATATAGGAGTTTTATCTAAGGGAACTCC	180
D _b	883	ATGAGGGTCCCGAATAATTGATGCCCATCGTAGGAGTCTTCTTAAGGGAACCCCC	942
Q _y	181	ACCTTCACTGCCACAACCCANATNGCCCGCAACTGCTATACTCTGCCACTCTTTGCATG	240
D _b	943	ACCTTCACTGCCACACCCATATGSCCCGCACTGCTATCATCTCTGCCACTCTTTGCATG	1002
Q _y	241	CATGCAAACTACTCATTTATTGGACAGGAAAATGATTAATCTCTAGTTGCTCTGGAGCACTT	300
D _b	1003	CATGCAAACTACTCATTTATTGGACAGGAAAATGATTAATCTCTAGTTGCTCTGGAGCACTT	1062
Q _y	301	GGAGCCACTGTCTGTGTTGGAATTACTTCAACCATACAGTATGCTGTGATGGGGTGGAAATT	360
D _b	1063	GGAGTCACTGTCTGTGTTGGAATTACTTCAACCAACTGGTATGCTGTGATGGGGTGGAGTT	1122
Q _y	361	CAAGGTCAAGCAAGAGAAAAACAAGTAAAGAAAGCAATCTCCCAACTGACCCGGGGAAT	420
D _b	1123	CAAGATCAGGACAGAGAAAAACAATGTAAGAAGTAATCTCCCACTCAACCGGGTACAT	1182
Q _y	421	AGCAACCCCTAGCCCTCAAAAGCACTAGTTCTCTCAAAACTCATGAAACCTTCGGTACC	480
D _b	1183	GGCACTCTAGCCCTCAAAAGCACTAGTCTCTCAAAACTCATGAAACCTTCGGTACC	1242
Q _y	481	CATACTCGCTGGTGAGCCTATTATATACCAACCTCACTCGGCTCCATCAGGTCTCAGCC	540
D _b	1243	CATACTCGCTGGTGAGCCTATTATATACCAACCTCACTCGGCTCCATCAGGTCTCAGCC	1302
Q _y	541	CAAAACCCCTACTAACTGTGATGTGCTCCCTCGCACTCAGGCGATACATTTCAATC	600
D _b	1303	CAAAACCCCTACTAACTGTGATGTGCTCCCTCGCACTCAGGCGATATGTTTCAATC	1362
Q _y	601	CCTGTTCTGTGAACATGGAAACACTTCAGACAGAAATAAACACCACTTCCGTTTTAGTA	660
D _b	1363	CCTGTACTGTGAACATGGAAACACTTCAGACAGAAATAAACACCACTTCCGTTTTAGTA	1422
Q _y	661	GGACCTCTGTGTTTCCAAATCTGGAATAAACCCATACTTCAAAACCTCACTGTGTAAATTT	720
D _b	1423	GGACCTCTGTGTTTCCAAATGTGGAATAAACCCATACCTCAAAACCTCACTGTGTAAATTT	1482
Q _y	721	AGCAATACTATAGACAAACAGCTCCCATGCAATCAGGTGGGTACACCTCCCAACCA	780
D _b	1483	AGCAATACTATACAAACCACTCCCAATGCATCAGGTGGGTAACTCTCCCAACAA	1542
Q _y	781	ATAGTCTGCCCTACCTCTCAGGAATATTTTTGTCTGTGTGTAACCTCAGCCTATCATTTGTTG	840
D _b	1543	ATAGTCTGCCCTACCTCTCAGGAATATTTTTGTCTGTGTGTAACCTCAGCCTATCATTTGTTG	1602
Q _y	841	AATGCTCTTCAGAAATCATGTGCTTCTCTCATTTAGTGGCCCTATGACCACTAC	900
D _b	1603	AATGCTCTTTCAGAAATCATGTGCTTCTCTCATTTAGTGGCCCTATGACCACTAC	1662
Q _y	901	ACTGAACAAGATTTTATACAATCATGTGCTTACTAAGCCCCCAACAAAGAGTACCCATT	960
D _b	1663	ACTGAACAAGATTTTATACAGTTATGTGCATATCTAAGCCCCCAACAAAGAGTACCCATT	1722
Q _y	961	CTTCTCTTTTGTATCAGACGAGGAGTGTAGGCHAGATAGTACTGCGATTTGGCAGTATC	1020
D _b	1723	CTTCTCTTTTGTATAGGACGAGGAGTGTAGGCHAGTACTGCGATTTGGCAGTATC	1782
Q _y	1021	AGAACTCTACTCAGTTCTTACTTACAAAATCTATCTCAAGAAAATAAATGGTGACATGAAACAG	1080
D _b	1783	AGAACTCTACTCAGTTCTTACTTACAAAATCTATCTCAAGAAAATAAATGGTGACATGAAACAG	1842
Q _y	1081	GTCACTGACTCCCTGGTGCACTTTGCAAGATCAACTTAACTCCCTAGCGAGTACTGTCCTTT	1140

Db	1943	GTCCCGACTCCCTGGTTCACCTTGGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT	1900
Qy	1141	CAAAATCGAAGAGCTTTAGACTTGTCTAAACGCGCAAAGAGGGGGAACCTGTTTATTTT	1200
Db	1903	CGAAATCGAAGAGCTTTAGACTTGTCTAAACGCGCTGAGAGAGGGGGAACCTGTTTATTTT	1962
Qy	1201	GGAGAAAGCGCTCTTATTATGTTTAAATCAATCCAGAAATGTCACTGAGAAAGTTAAAGAA	1260
Db	1963	GGGGAAGAAGTCTGTTTATTAATGTTTAAATCAATCCGGAATCGTCACTGAGAAAGTTGAAGAA	2022
Qy	1261	ATTCCGAGATCGAATACAAATCTAGACGACAGAGAGCTTCAAAACACCGAACCTGGGGCTC	1320
Db	2023	ATTCCAGATCGAATACAACTATAGCAGAGAGAGCTTCGAAACACACTGGACCTGGGGCTC	2082
Qy	1321	CTACGCAATGAGATCGCCCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAAATATTG	1380
Db	2083	CTACGCGGATGAGTGGCCCTGGATCTCTCCCTTTTAGGACCTCTAGCAGCTATAATATTG	2142
Qy	1381	TTACTCTCTTTTGGACCTGTATCTTTAACTCTCTTCTTAAAGTTTGTCTCTTCAGAAATT	1440
Db	2143	CTACTCTCTTTTGGACCTGTATCTTTGACCTCTTGTAACTTGTCTCTTCCAGAAATC	2202
Qy	1441	GAAGCTGTAAAGCTACAGATGGTCTTACAATGGAAACCCCA	1481
Db	2203	GAAGCTGTGAAACTACAAATGGAGCCCAAGATGACGTCCAA	2243
RESULT 21			
AAAX25661			
ID	AAAX25661 standard; cDNA to mRNA; 2782 BP.		
AC	AAAX25661;		
XX			
DT	21-MAY-1999 (first entry)		
XX			
DE	Human endogenous retrovirus W clone cl.PH74.		
XX			
KW	Clone; human endogenous retrovirus; genome; autoimmune disease; multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes; disseminated lupus erythematosus; pregnancy; chromosomal marker; SS.		
XX			
OS	Human endogenous retrovirus.		
XX			
PN	WO9902696-Al.		
XX			
PD	21-JAN-1999.		
XX			
PF	06-JUL-1998; 98WO-FR001442.		
XX			
PR	07-JUL-1997; 97FR-00008815.		
XX			
PA	(INMR) BIO MERIEUX.		
XX			
PI	Beseme F, Blond J, Bouton O, Mandrand B, Mallet F;		
XX			
DR	WPI; 1999-120897/10.		
XX			
PT	New nucleic acid sequences from human endogenous retrovirus-W - expressed exclusively in placenta and useful in diagnosis and therapy of autoimmune disease, and abnormal or failed pregnancy.		
XX			
PS	Claim 1; Page 60-63; 106pp; French.		
XX			
CC	This sequence represents clone cl.PH74 of the human endogenous retrovirus (HERV) W genome. The nucleic acids, their fragments or peptides encoded by them are markers of autoimmune disease (e.g. multiple sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-dependent diabetes and related pathologies) and of abnormal or unsuccessful pregnancy and can be used as chromosomal markers for susceptibility to these conditions, or proximity markers of genes associated with this susceptibility		
XX			
SQ	Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 U; 0 Other;		

Query Match									
Best Local Similarity 88.28; Score 1306.6; DB 2; Length 2782;									
Matches 1372; Conservative 0; Mismatches 109; Indels 0; Gaps 0;									
QY	1	ATGGCCCTCCCTTATCATCTTTCTCTTTACTTGTCTCTTACCTCCCTTTTCGCTCTCACT	60						
DB	763	ATGGGCTCCCTTATCATATTTTCTCTGTAGTGTCTTTCACCCCTGTTTCACTCTCACT	822						
QY	61	GCACCCCTCCATCTGCTCTGACCAACAGTAGTCTCCCTTACCAAGAGTTTCTATGAAGA	120						
DB	823	GCACCCCTCCATGCGCTGTATGACCAAGTAGTCTCCCTCACCCAGAGTTTCTATGAGA	882						
QY	121	ACGGGCTTCTGGAATATTTGATGCCCAATCATATAGGAGTTTATCTAAGGGAATCTC	180						
DB	883	ATGACGGCTCCCGGAATATTTGATGCCCCATCGTATAGGAGTCTTCTAAGGGAACCC	942						
QY	181	ACCTTCACTGCCACACCCCATATGCCCCGCACTGCTATTAATCTGTGCACTCTTTGATG	240						
DB	943	ACCTTCACTGCCACACCCCATATGCCCCGCACTGCTATTAATCTGTGCACTCTTTGATG	1002						
QY	241	CATGCAAACTACTCATTTATGGACAGGGAATGATTAATCTTAGTTGTCTGGAGGACTT	300						
DB	1003	CATGCAAACTACTCATTTATGGACAGGGAATGATTAATCTTAGTTGTCTGGAGGACTT	1062						
QY	301	GGAGCCTCTGTCTGTGGACTTACTTCAACCATACCAAGTATGCTGATGGGGGTGGAAT	360						
DB	1063	GGAGTCACTGTCTGTGGACTTACTTCAACCATACCAAGTATGCTGATGGGGGTGGAAT	1122						
QY	361	CAAGGTCAGGCAAGAGAAACAAAGTAAGGAAGCAATCTCCCACTGACCCCGGGGACAT	420						
DB	1123	CAAGATCAGGCAAGAGAAACAAAGTAAGGAAGCAATCTCCCACTGACCCCGGGGACAT	1182						
QY	421	AGCACCTCTAGCCCTACAAAGGAGTAGTTCTCTCAAACTACATGAAACCTCCCGTACC	480						
DB	1183	GGCACCTCTAGCCCTACAAAGGAGTAGATCTCTCAAACTACATGAAACCTCCCGTACC	1242						
QY	481	CATATCGCTGTGTGAGCCTATTTAATACACCCCTCACTCGGCTCCATGAGGTCAGCC	540						
DB	1243	CATATCGCTGTGTGAGCCTATTTAATACACCCCTCACTCGGCTCCATGAGGTCAGCC	1302						
QY	541	CAAAACCTCTACTACTGTGGATGTCCTCCCTGCACTTCAAGGCAATATTTCAATC	600						
DB	1303	CAAAACCTCTACTACTGTGGATGTCCTCCCTGCACTTCAAGGCAATATTTCAATC	1362						
QY	601	CCTGTTCTCTGAACTGGAACAACTTCAGCACAGAAATAAACCACTTCCGTTTGTAGTA	660						
DB	1363	CCTGTTCTCTGAACTGGAACAACTTCAGCACAGAAATAAACCACTTCCGTTTGTAGTA	1422						
QY	661	GGACCTCTTGTTCCTCAATCTGGAATAAACCCATACCTCAAACTCACTGTGTAAATTT	720						
DB	1423	GGACCTCTTGTTCCTCAATCTGGAATAAACCCATACCTCAAACTCACTGTGTAAATTT	1482						
QY	721	AGCAATCTATAGACACCAACAGCTCCCAATGATCAGGTGGGTAAACCTCCACACGA	780						
DB	1483	AGCAATCTATAGACACCAACAGCTCCCAATGATCAGGTGGGTAAACCTCCACACAA	1542						
QY	781	ATAGTCTGCTTACCTCAGGAATATTTTGTCTGTGTACCTCAGCTATCATTTGTTTG	840						
DB	1543	ATAGTCTGCTTACCTCAGGAATATTTTGTCTGTGTACCTCAGCTATCATTTGTTTG	1602						
QY	841	AATGGCTCTTTCAGAACTATGTGCTCTCTCTCAATTTAGTGCCCTCATGACCATCTAC	900						
DB	1603	AATGGCTCTTTCAGAACTATGTGCTCTCTCTCAATTTAGTGCCCTCATGACCATCTAC	1662						
QY	901	ACTGAACAAGATTTATACATCATGTGTCATCTAAGCCCCACAAACAAAGAGTACCCATT	960						
DB	1663	ACTGAACAAGATTTATACATCATGTGTCATCTAAGCCCCACAAACAAAGAGTACCCATT	1722						
QY	961	CTTCTCTTCTGATACAGACAGAGTGCTAGGACAGCTAGTGTGCTGCACTGAGTATC	1020						
DB	1723	CTTCTCTTCTGATACAGACAGAGTGCTAGGACAGCTAGTGTGCTGCACTGAGTATC	1782						

QY	1021	ACAACTCTACTCAGTTCTACTACAACTATCTCAAGAAATTAATGGTGACATGGACAG	1080
DB	1783	ACAACTCTACTCAGTTCTACTACAACTATCTCAAGAAATTAATGGTGACATGGACAG	1842
QY	1081	GTCACTGACTCCCTGCTGCTACCTTGCAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT	1140
DB	1843	GTGCGCGACTCCCTGCTGCTACCTTGCAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTT	1902
QY	1141	CAAAATCGAAGAGCTTTAGACTTTGCTTAAACCGCCAAAGAGGGGGAACCTGTTATTTTA	1200
DB	1903	CGAAATCGAAGAGCTTTAGACTTTGCTTAAACCGCTGAGAGAGGGGGAACCTGTTATTTTA	1962
QY	1201	GGAGAGAAACGCTGTTTATTTATGTTAATCAATCCAGATTTGCTACTGAGAAAGTTAAAGAA	1260
DB	1963	GGGAGAGAAAGCTGTTTATTTATGTTAATCAATCCAGATTTGCTACTGAGAAAGTTAAAGAA	2022
QY	1261	ATTTCAGATCGAATACAAATGTAGCAGAGGAGCTTCAAAACACCCGAAACCTGGGGCTC	1320
DB	2023	ATTTCAGATCGAATACAAATGTAGCAGAGGAGCTTCAAAACACCTGGGGCTC	2082
QY	1321	CTCAGCAATGGAATGCGCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG	1380
DB	2083	CTCAGCGATGGAATGCGCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG	2142
QY	1381	TTACTCTCTTTGGACCTGTATCTTTAACTCCCTTGTAAAGTTTCTCTTCCAGAAAT	1440
DB	2143	CTACTCTCTTTGGACCTGTATCTTTGACCTCTTGTAACTTGTCTCTCCAGAAATC	2202
QY	1441	GAACTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA	1481
DB	2203	GAACTGTAAAGCTACAAATGGAACCCCAAGATGCACTCCAA	2243

RESULT 22
AAA59211
ID AAA59211 standard; DNA; 2782 BP.

XX AAA59211;

DT 07-NOV-2000 (first entry)

XX 5' non coding, 3' pol gene, and 3' non coding sequences of HERV-W.

XX Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
KW gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.

XX Homo sapiens.

XX WO200043521-A2.

XX 27-JUL-2000.

XX 21-JAN-2000; 2000WO-FR000144.

XX 21-JAN-1999; 99FR-00000888.

XX (INMR) BIO MERIEUX.

XX Paranhos-Baccala G, Mallet F, Voisset C;

XX WPI; 2000-499229/44.

XX New nucleic acid from human endogenous retrovirus, useful e.g. for
PT diagnosis of autoimmune disease and complications of pregnancy, contains
PT at least part of the gag gene.

XX Disclosure; Page 46-47; 53pp; French.

XX The present sequence represents an endogenous retroviral nucleic acid
CC fragment, which is associated with an autoimmune disease, and is
CC integrated into the human genome. The fragment is originally derived from
CC a novel retrovirus, human endogenous retrovirus W (HERV-W). The HERV-W
CC retrovirus is associated with autoimmune disease, failure of pregnancy or

PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
XX WPI; 2003-569235/53.
XX
XX New polynucleotides, useful for expressing recombinant proteins for
PT analysis, characterization or therapeutic use, or as markers for tissues
PT in which the corresponding protein is preferentially expressed.
XX
XX Disclosure; SEQ ID NO 2131; 1177pp; English.
XX
XX The invention comprises the amino acid and coding sequences of novel
CC proteins. The DNA and protein sequences of the invention are useful as:
CC markers for tissues in which the corresponding protein is preferentially
CC expressed; as molecular weight markers on gels; as chromosome markers or
CC tags; to identify chromosomes or to map related gene positions; and to
CC compare with endogenous DNA sequences in patients to identify potential
CC genetic disorders. The present DNA sequence was used in the
CC exemplification of the invention.
XX
SQ Sequence 3464 BP; 1038 A; 913 C; 666 G; 845 T; 0 U; 2 Other;
Query Match 87.7%; Score 1298.2; DB 10; Length 3464;
Best Local Similarity 93.0%; Pred. No. 0;
Matches 1383; Conservative 0; Mismatches 98; Indels 6; Gaps 2;
QY 1 ATGGCCCTCCCTTATCATATCTTTCTTTACTTGTCTCTTACCCCTTTTCGCTCTCACT 60
DB 1430 ATGGCCCTCCCTTATCATATCTTTCTTTACTTGTCTTTTACCCCTTTCTCACTCTCACT 1489
QY 61 GCACCCCTCCATGCTGCTCAACACAGTAGTCTCCCTTACCAAGATTTCTATGAAGA 120
DB 1490 GCACCCCTCCATGCTGCTCAACACAGTAGTCTCCCTTACCAAGATTTCTATGAAGA 1549
QY 121 ACCGGCTTCTCGAATATTGATGCCCATATATAGGAGTTTACTAAGGGAATCTC 180
DB 1550 ATGCAGCGTCCCGGAATATTGATGCCCATATATAGGAGTTTCTTAAGGGAATCTC 1609
QY 181 ACCTTCACTCCCAACCCATATGCCCCGCAACTGTCTATTAACCTCTGCACTCTTTGCATG 240
DB 1610 ACCTTCACTCCCAACCCATATGCCCCGCAACTGTCTATTAACCTCTGCACTCTTTGCATG 1669
QY 241 CATGCAAAATCTCATTTATGGACAGGGAATGATTAATCCTAGTTGTCTCTGAGAGACTT 300
DB 1670 CATGCAAAATCTCATTTATGGACAGGGAATGATTAATCCTAGTTGTCTCTGAGAGACTT 1729
QY 301 GGAGCCACTGTCTGTGGACTTACTTCAACCATACCAGTATGCTGATGGGGTGAATT 360
DB 1730 GGAGTCACTGTCTGTGGACTTACTTCAACCATACCAGTATGCTGATGGGGTGAATT 1789
QY 361 CAAGGTTCAGGCAAGAGAAAAAAGTAAGGAAGCAATCTCCCACTGACCCGGGGACAT 420
DB 1790 CAAGTTCAGGCAAGAGAAAAAAGTAAGGAAGCAATCTCCCACTGACCCGGGGACAT 1849
QY 421 AGCACCCCTAGCCCTTACAAAGGACTAGTTCTCTCAAAATACATGAACCCCTCCGTAAC 480
DB 1850 AGCACCCCTAGCCCTTACAAAGGACTAGTTCTCTCAAAATACATGAACCCCTCCGTAAC 1909
QY 481 CATACTCGCTGTGAGCCCTATTATTAACACCCCTCACTCGGCTCCATGAGGCTCAGCC 540
DB 1910 CATACTCGCTGTGAGCCCTATTATTAACACCCCTCACTCGGCTCCATGAGGCTCAGCC 1969
QY 541 CAAACCCCTACTAATCTGTGATGTGCTCCCTCGCTGACCTTCAAGGCCATACATTTCATC 600
DB 1970 CAAACCCCTACTAATCTGTGATGTGCTCCCTCGCTGACCTTCAAGGCCATATGTTCAATC 2029
QY 601 CCTGTTCTGAAACAAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTGTAGTA 660
DB 2030 CCTGTACTGACAAATGGAACAACTTCAGCACAGAAATAAACACCACTTCCGTTTGTAGTA 2089
QY 661 GGACCTCTGTTTCCAAATGGAATAAACCACTCAACCTCAACCTCAACCTGTTGAAAAATTT 720

DB 2090 GGACCTCTGTTTCCAAATCTGGAATAATACCCATACCTCAACCTCACTGTGTAAAAATTT 2149
QY 721 AGCAATACTATAGACACAAACCAAGCTCCCAATAGCATCAGGTGGGTAAACCTCCACACAGA 780
DB 2150 AGCAATACTATAGACACAAACCAAGCTCCCAATAGCATCAGGTGGGTAAATCTCTCCACACAA 2209
QY 781 ATAGTCTGCTACCTACCTCAGGAATATTTTGTCTGTGTGACCTCAGCCTATCATTTGTTG 840
DB 2210 ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGTGACCTCAGCCTATCATTTGTTGTTG 2269
QY 841 AATGGCTCTTCAGAAATCTATGTCTCTCTCTCAATTTAGTGTGCCCTATGACCATCTAC 900
DB 2270 AATGGCTCTTCAGAAATCTATGTCTCTCTCTCAATTTAGTGTGCCCTATGACCATCTAC 2329
QY 901 ACTGAACAAGATTATACAAATCATGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 956
DB 2330 ACTGAACAAGATTATACAAATCATGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2389
QY 957 CATTTCTTCTTTTGTATCAGAGCAGGAGTGTAGGACACTAGGTAGTCTGCGATTTGCGAG 1016
DB 2390 CATTTCTTCTTTTGTATTTGGAGCAGGAGTACTAGGCGGACTAGGTACTGGCAATTTGCGG 2449
QY 1017 TATCAACAACCTCTACTCAGTTCTTACTACAAACTATCTCAAGAAATAAATGGTGACATGGA 1076
DB 2450 TATTACAACTCTACTCAGTTCTTACTACCAAACTATCTCAAGAAATAAATGGTGACATGGA 2509
QY 1077 ACAGGT--CACTGACTCCCTGGTCACTTGTCAAGATCAACTTAACTCCCTTAGCAGAGTA 1134
DB 2510 ACAGGTGCGCCGACTCCCTGGTCACTTGTCAAGATCAACTTAACTCCCTTAGCAGAGTA 2569
QY 1135 GTCTTCAAAATCGAAGAGCTTTAGACTTGTCAACCGCAAAAGAGGGGAACTGTTTA 1194
DB 2570 GTCTTCAAAATCGAAGAGCTTTAGACTTGTCAACCGCTGAAAGAGGGGAACTGTTTA 2629
QY 1195 TTTTATGAGGAAGAAGCGCTTATTATTTATTTAATCAATCCAGAAATTTGCTACTGAGAAAGTT 1254
DB 2630 CTTTATGAGGAAGAAGCGCTTATTATTTAATCAATCCAGAAATTTGCTACTGAGAAAGTT 2689
QY 1255 AAAGAAATTCGAGATCGAATACAAATGTAGACAGAGAGCTTCAAAACACCGAACGCTGG 1314
DB 2690 AAAGAAATTCGAGATCGAATACAAATGTAGACAGAGAGCTTCAAAACACCGAACGCTGG 2749
QY 1315 GGCCTCTCAGCAATGGATGCCCTGGTCTCCCTTCTTAGGACCTCTAGCAGCTCTA 1374
DB 2750 GGCCTCTCAGCAATGGATGCCCTGGTCTCCCTTCTTAGGACCTCTAGCAGCTCTA 2809
QY 1375 ATATTGTTACTCTCTTTGGACCTGTATCTTTAACTCTCTTAAAGTTTGTCTCTTCC 1434
DB 2810 ATATTGTTACTCTCTTTGGACCTGTATCTTTAACTCTCTTAAAGTTTGTCTCTTCC 2869
QY 1435 AGAATTTGAAGCTGTAAAGCTTACAGATGGTCTTTTAAATGGAACCCCA 1481
DB 2870 AGAATCGAAGCTGTAAAGCTTACAGATGGGAGCCCAAGATGAGTCCCA 2916
RESULT 24
ADF59718
ID ADF59718 standard; cDNA; 9502 BP.
XX
AC ADF59718;
XX
DT 12-FEB-2004 (first entry)
XX
DE Human contig polynucleotide sequence SEQ ID NO:2085.
XX
KW biological activity; genetic engineering; hybridisation probe; oligomer;
KW primer; chromosome mapping; gene mapping; recombinant protein production;
XX
OS human; gene; ss.
XX
XX Homo sapiens.
XX
XX WO2003080795-A2.
XX

PD 02-OCT-2003.
XX
PF 09-AUG-2002; 2002WO-US025485.
XX
PR 09-AUG-2001; 2001US-0311261P.
XX (HYSE-) HYSEQ INC.
PA
XX
PI Tang YT, Yang Y, Wang Z, Weng G, Ma Y;
XX WPI; 2003-876918/81.
XX P-PSDB; ADF60170.
DR
DR
XX New polynucleotides, useful as hybridization probes, oligomers or
PT primers, for chromosome or gene mapping, for the recombinant production
PT of proteins, and for generating antisense DNA or RNA.
XX
XX Example 2; SEQ ID NO 2085; 571pp; English.
XX
XX The present invention describes isolated polynucleotide sequences (I),
CC which encode polypeptides (II) with biological activity. Also described:
CC (1) a vector comprising (i); (2) an expression vector comprising (i); (3)
CC a host cell genetically engineered to comprise (i) which is operatively
CC associated with a regulatory sequence that modulates expression of (i) in
CC the host cell; (4) a polypeptide (ii) encoded by (i); (5) a composition
CC comprising the polypeptide of (4) and a carrier; (6) an antibody directed
CC against the polypeptide of (4); (7) detecting (i) or the polypeptide of
CC (4) in a sample; (8) identifying a compound that binds to the polypeptide
CC of (4); (9) producing the polypeptide of (4); and (10) a collection of
CC polynucleotides comprising at least one of the polynucleotide sequences
CC (i). The polynucleotides (i) can be used as hybridisation probes,
CC oligomers or primers, for chromosome or gene mapping, for the recombinant
CC production of proteins, and for generating antisense DNA or RNA. The
CC present sequence represents a human contig polynucleotide sequence, which
XX is used in an example from the present invention.
XX
SQ Sequence 9502 BP; 2813 A; 2433 C; 1992 G; 2263 T; 0 U; 1 Other;
Query Match 87.7%; Score 1298.2; DB 10; Length 9502;
Best Local Similarity 93.0%; Pred. No. 0;
Matches 1383; Conservative 0; Mismatches 98; Indels 6; Gaps 2;
QY 1 ATGGCCCTCCCTATCACTATTTCTCTTTACTGTTCTTTACCCCTTTGCTCTCACT 60
Db 1430 ATGGCCCTCCCTATCACTATTTCTCTTTACTGTTCTTTACCCCTTTGCTCTCACT 1489
QY 61 GCACCCCTCAGTGTCTGACACAGTAGTCCCTTACCAAGAGTTTCTATGAAGA 120
Db 1490 GCACCCCTCAGTGTCTGACACAGTAGTCCCTTACCAAGAGTTTCTATGAGA 1549
QY 121 ACGCGGCTTCTGGAATATTGATGCCCATCATATAGAGTTTATCTAAGGGAACTCC 180
Db 1550 ATGCAGCTCCGGAATATTGATGCCCATCATATAGAGTTTCTTACGGGAACCC 1609
QY 181 ACCTTCACTGCCACACCCATATGCCCCGCACTGCTATTAATCTGCTTGGAGCTT 240
Db 1610 ACCTTCACTGCCACACCCATATGCCCCGCACTGCTATTAATCTGCTTGGAGCTT 1669
QY 241 CATGCAATATCACTATTGACAGGAAATGATTAACTTCTGCTTGGAGCTT 300
Db 1670 CATGCAATATCACTATTGACAGGAAATGATTAACTTCTGCTTGGAGCTT 1729
QY 301 GGAGCCACTGTCTGTGGACTTACTTCCACCATACAGTATGTCTGATGGGGTGGAAATT 360
Db 1730 GGAGTCACTGTCTGTGGACTTACTTCCACCATACAGTATGTCTGATGGGGTGGAGTT 1789
QY 361 CAAGGTCAAGCAAGAAACCAAGTAAGAGCAATCTCCCACTGACCCGGGACAT 420
Db 1790 CAAGATCAGCAAGCAAGAAACCAAGTAAGAGCAATCTCCCACTGACCCGGGACAT 1849
QY 421 AGCACCCCTAGCCCTTACAAAGGACTAGTTCTCTCAAAACTACATGAACCCCTCCGTACC 480
Db 1850 AGCACCCCTAGCCCTTACAAAGGACTAGTTCTCTCAAAACTACATGAACCCCTCCGTACC 1909

QY 481 CATACTCGCTGGTGAAGCCTATTATACCAACCTCTACTCGGCTCCATGAGGTCTCAGCC 540
Db 1910 CATACTCGCTGGTGAAGCCTATTATATACCAACCTCTACTCGGCTCCATGAGGTCTCAGCC 1969
QY 541 CAAAACCTCTACTGTTGATGCTCCCTCGCACTTCAGGCCATACATTTCAATC 600
Db 1970 CAAAACCTCTACTGTTGATGCTCCCTCGCACTTCAGGCCATATGTTCAATC 2029
QY 601 CCTGTCTCTGAACCAATGGAACAACTTCAGCACAGAAATAAACACCACTTCGTTTATGTA 660
Db 2030 CCTGTCTCTGAACCAATGGAACAACTTCAGCACAGAAATAAACACCACTTCGTTTATGTA 2089
QY 661 GGACCTCTTGTTCATATCGAAATAACCACTCAACCTCAACCTCAACCTCAACCTCAACCT 720
Db 2090 GGACCTCTTGTTCATATCGAAATAACCACTCAACCTCAACCTCAACCTCAACCTCAACCT 2149
QY 721 AGCAATATATAGACACAAACCACTCCCAATGCAATCAGTGGGTAAACACCTCCACACGA 780
Db 2150 AGCAATATATAGACACAAACCACTCCCAATGCAATCAGTGGGTAAACCTCCACACGA 2209
QY 781 ATAGTCTGCTACCTCCCTCAGGAATATTTTGTCTGTGTGCTCCTCAGCCATCATGTTTG 840
Db 2210 ATAGTCTGCTACCTCCCTCAGGAATATTTTGTCTGTGTGCTCCTCAGCCATCATGTTTG 2269
QY 841 AATGGCTCTTCAGATCTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
Db 2270 AATGGCTCTTCAGATCTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2329
QY 901 ACTGAACCAAGATTTATACAACTCATGTC---GTACTTAAGCCCAACAAACCAAGATGACC 956
Db 2330 ACTGAACCAAGATTTATACAGTTATGTAATCTCTAAAGCCCAACAAACCAAGATGACC 2389
QY 957 CATTTCTCTTTTGTATCAGAGAGGAGTGTCTAGCAGACTAGGTACTGGCATTTGGCAG 1016
Db 2390 CATTTCTCTTTTGTATTTGGAGAGGAGTACTAGCGGAGTACTAGGTACTGGCATTTGGCAG 2449
QY 1017 TATCAACCTCTACTCTAGTTCTACTACAACTATCTCAAGAAATAAATGTTGACATGGA 1076
Db 2450 TATCAACCTCTACTCTAGTTCTACTACAACTATCTCAAGAAATAAATGTTGACATGGA 2509
QY 1077 ACAGGT--CACTGACTCCCTGTGCTCACTTGAAGATCAACTTAACCTCCCTAGCAGCTA 1134
Db 2510 ACAGGTGCGCCGACTCCCTGTGCTCACTTGAAGATCAACTTAACCTCCCTAGCAGCTA 2569
QY 1135 GTCTTCAAAATCGAAGAGCTTTAGACTTGTAAACCGCAAGAGGGGAAACCTGTTTAA 1194
Db 2570 GTCTTCAAAATCGAAGAGCTTTAGACTTGTAAACCGCTGAAGAGGGGAAACCTGTTTAA 2629
QY 1195 TTTTATAGGAGAAAGAGCTGTTTATTTATTAATCAATCCAGAAATGTCCTAGAGAAAGTT 1254
Db 2630 CTTTATAGGAGAAAGAGCTGTTTATTTATTAATCAATCCAGAAATGTCCTAGAGAAAGTT 2689
QY 1255 AAAGAAATTCAGATCGAATAAATGTAGACAGAGAGCTTCAAACACCAAGAGCTGG 1314
Db 2690 AAAGAAATTCAGATCGAATAAATGTAGACAGAGAGCTTCGAAACCACTGGACCCCTGG 2749
QY 1315 GGCTCTCTAGCAAAAGAGTGGCTTCTCCCTCTCTAGGAGCTCTAGCAGCTCTA 1374
Db 2750 GGCTCTCTAGCAAAAGAGTGGCTTCTCCCTCTCTAGGAGCTCTAGCAGCTCTA 2809
QY 1375 ATATTGTTACTCTCTTTGGACCTGTATCTTTTAACTCTCTGTTTAAAGTTTGTCTTCC 1434
Db 2810 ATATTGTTACTCTCTTTGGACCTGTATCTTTTAACTCTCTGTTTAAAGTTTGTCTTCC 2869
QY 1435 AGAATTGAAGCTGAAGCTACAGATGCTTACAAATGGAACCCCA 1481
Db 2870 AGAATTGAAGCTGAAGCTACAGATGCTTACAAATGGAAGCCCAAGATGAGTCCAA 2916

RESULT 25
ABN97978
ID ABN97978 standard; DNA; 46340 BP.

[illegible]

Db	33234	CTTCTTAGGACTCTAGCAGCTCTAATATTGATCTCTCTTTGGACCTCTATCTTTAA	33293
Qy	1410	CTCTCTTGTAAAGTTTGTCTCTTTCCAGAAATGAAGCTGTAAGCTTCTTACA	1469
Db	33294	CTCTCTTGTAAAGTTTGTCTCTTTCCAGAAATGAAGCTGTAAGCTTCTTACA	33353
Qy	1470	AATGGAACCCCA	1481
Db	33354	AATGGAACCCCA	33365
RESULT 26			
ABN97931/c			
ID ABN97931 standard; DNA; 1799 BP.			
XX	AC	ABN97931;	
XX	XX	01-AUG-2002 (first entry)	
XX	DE	Human retroviral sequence HE3.	
XX	XX	Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;	
KW	KW	multiple sclerosis; ds.	
XX	XX	Human endogenous retrovirus.	
OS	XX	WO9967395-A1.	
XX	XX	29-DEC-1999.	
XX	XX	23-JUN-1999; 99WO-FR001513.	
XX	XX	23-JUN-1998; 98FR-00007920.	
XX	PA	(INRM) INSERM INST NAT SANTE & RECH MEDICALE.	
XX	PI	Alliel PM, Perin J, Rieger F;	
XX	XX	WPI; 2000-160587/14.	
XX	XX	New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used	
PT	PT	for diagnosis, treatment and prevention of autoimmune and neurological	
PT	PT	diseases.	
XX	PS	Claim 3; Page 134; 225pp; French.	
XX	CC	The present invention relates to new nucleic acid sequences of human	
CC	CC	endogenous retrovirus, HERV-7q, which is located on chromosome 7q.	
CC	CC	Regulatory elements associated with HERV-7q may alter expression of other	
CC	CC	genes (even remote genes) on the same chromosome, inducing immunological	
CC	CC	and/or neurological changes (which may be pathological or protective/	
CC	CC	curative). HERV-7q peptides can be used to improve efficiency of the	
CC	CC	immune response, e.g. in immunotherapy. HERV-7q peptides and their coding	
CC	CC	sequences can be used in immunogenic or vaccinating compositions, for	
CC	CC	protection against autoimmune diseases, particularly multiple sclerosis.	
CC	CC	The peptides may also be used (by sequence comparison) to detect/identify	
CC	CC	endogenous retroviruses that are abnormally expressed in cancer,	
CC	CC	neuropathologies or other autoimmune diseases. The present sequence was	
XX	SQ	used to illustrate the invention	
Sequence 1799 BP; 488 A; 331 C; 482 G; 498 T; 0 U; 0 Other;			
Query Watch 86.3%; Score 1278.8; DB 3; Length 1799;			
Best Local Similarity 92.1%; Pred. No. 0;			
Matches 1393; Conservative 0; Mismatches 87; Indels 32; Gaps 3;			
Qy	1	ATGGCCCTCCCTATACATATTTCTTTTACTGTCTTTACCCCTTTGCTCTCACT	60
Db	1762	ATGGCCCTCCCTATACATATTTCTTTTACTGTCTTTACCCCTTTTCACTCTCACT	1703
Qy	61	GCACCCCTCCATGCTGCTGTAACACAGTAGTCCCTTACCAAGAGTTTCTATGAAGA	120
Db	1702	GCACCCCTCCATGCTGCTGTAACACAGTAGTCCCTTACCAAGAGTTTCTATGAAGA	1643

Qy	121	ACGGGCTTCTCGAATATATTGATGCCCCCATATAGGAGTTTATCTAAGGAAACTCC	180
Db	1642	ATGGGCTTCCAGAAATATTGATGCCCCCATCAATAGGAGTTTACCTAAGGAAACTCC	1583
Qy	181	ACCTTCACTGCGCCACACCATATGCCCCGCAATGCTATTAACCTCTGCGCATCTCTTTGCATG	240
Db	1582	ACCTTCACTGCGCCACACCATATGCCCCCAACATGCTATTAACCTCTGCGCATCTCTTTGCATG	1523
Qy	241	CATGCAATACTCTATTATTGGACAGGAAATGATTAATCTTAGTTGCTCTGGAGGACTT	300
Db	1522	CATGCAATACTCTATTATTGGACAGGAAATGATTAATCTTAGTTGCTCTGGAGGACTT	1463
Qy	301	GGAGCCACTGTCTGTGGACTTACTTCAACCATACAGTATGCTGATGGGGTGGAAATT	360
Db	1462	GGAGCCACTGTCTGTGGACTTACTTCAACCATACAGTATGCTGATGGGGTGGAGATT	1403
Qy	361	CAAGTCAAGCAAGAGAAAAAAGTAAGGAAGCAATCTCCCAACTGACCCCGGGGACAT	420
Db	1402	CAAGATCAGGCAAGAGAAAAACATGTAAGGAAGTAACCTTCCCAACTGACCCCGGTACAT	1343
Qy	421	AGCACCCCTAGCCCCCTACAAAGGACTAGTTCTCTCAAAATACATGAAGAACCTCTCGTACC	480
Db	1342	AGCACCCCTAGCCCCCTACAAAGGACTAGATCTCTTAAACTACATGAAGAACCTCCATACC	1293
Qy	481	CATACCTCGCTGGTGAAGCTTATTTAATACACCTCTCACTCGGCTCCATGAGTCTCGACC	540
Db	1282	CATACCTCGCTGGTGAAGCTTATTTAATACACCTCTCACTCGGCTCCATGAGTCTCGGTC	1223
Qy	541	CAAAACCTCTACTTAACTGTGGATGCTCCCTCCCTGCACTTTCAGGCCATATCAATTCAATC	600
Db	1222	CAAAACCTCTACTTAACTGTGGATGCTCCCTCCCTGCACTTTCAGGCCATATCAATTCAATC	1163
Qy	601	CCTGTTCTCTGAACAATGGAACAACCTTCAGACAGAAATAAACACACACTTCGTTTACTGA	660
Db	1162	CCTGTTCTCTGAACAATGGAACAACCTTCAGACAGAAATAAACACACACTTCGTTTACTGA	1103
Qy	661	GGACCTCTGTTTCCAACTGGAATAACCCATCTCAAACTCTCACTGTGTAAATTT	720
Db	1102	GGACCTCTGTTTCCAACTGGAATAACCCATCTCAAACTCTCACTGTGTAAATTT	1043
Qy	721	AGCAATCTATAGACACAACCAAGCTCCCAATGCATCAGGTGGGTAAACCTCCACACGA	780
Db	1042	AGCAATCTATAGACACAACCAAGCTCCCAATGCATCAGGTGGGTAAACCTCCACACGA	983
Qy	781	ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGTGTACCTCAGCTCTATCATTTGTTG	840
Db	982	ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGTGTACCTTACCTTATCGTTGTTG	923
Qy	841	AATGGCTCTTCAGAACTATGTGCTTCTCTCATTTCTTAGTCCCTTATGACCATCTAC	900
Db	922	AATGGCTCTTCAGAACTATGTGCTTCTCTCATTTCTTAGTCCCTTATGACCATCTAC	864
Qy	901	ACTGAAACAAGATTATTAACAATCATGTCGTACCTTAAGCCCAACAACAAAGAGTACCAATT	960
Db	863	ACTGAAACAAGATTATTAACAATCATGTCGTACCTTAAGCCCAACAACAAAGAGTACCAATT	804
Qy	961	CTTCTCTTTGTTATCAGACGAGGCTGTAGGACAGACTAGGTACTGGCATTTGGCAGTATC	1020
Db	803	CTTCTCTTTGTTATCGGAGCAGGAGTGTAGGTGGAGTCTAGGTCTGGCATTTGGCGTACC	744
Qy	1021	ACAACTCTACTCAGTTTCTACTACAAATATCTCAAGAAATAAATGGTGTGATGGAAACAG	1080
Db	743	ACAACTCTACTCAGTTTCTACTACAAATATCTCAAGAAATAAATGGTGTGATGGAAATGG	684
Qy	1081	GTCACTCACTCCCTGGTCACTTGTGCAAGATCACTTAACCTCCCTAGCAGAGTAGTCTCT	1140
Db	683	GTTCGCCACTCCCTGGTCACTTGTGCAAGATCACTTAACCTCCCTAGCAGAGTAGTCTCT	624
Qy	1141	CAAAATCGAGAGCTTTAGACTTGTCTAAACCCCAAGAGGGGGAACCTCTTTATTTTAA	1200
Db	623	CAAAATCGAGAGCTTTAGACTTGTCTAAACCCCAAGAGGGGGAACCTCTTTATTTTAA	564

Qy	1201	GGAGAAAGCGCTG-----TTATTATGTTAATCAA	1230
Db	563	GGGGAAGAATGTTGTTATATGTTATTTAGCGGAAGAATGCTGTTATATGTTAATCAA	504
Qy	1231	TCCAGAAATGTCACTGAGAAAAGTTAAAGAAAATTCGAGATCGAATACAAATGTAGAGCAGAG	1290
Db	503	TCCTGAATTTGTCA CAGAGAAAAGTTGAAGAAAATTCGAGATTGAATACAAACGTAGAACAGAG	444
Qy	1291	GAGCTTC-AAAACACCGAAGCGCTGGGGCCCTCCTCAGCCAAATGGATGCCCTGGGTTCTCC	1349
Db	443	GAGCTTCAAAAACACACAGACCTGGGGCCCTCCTCAGCCAAATGGATGCCCTGGATTTCTCCC	384
Qy	1350	CTTCTTAGGACCTTAGCAGCTCTAAATATTGTTACTCCTCTTTGGACCCCTGTATCTTTAA	1409
Db	383	CTTCTTAGGATCTCTAGCAGCTCTAAATATTGATACTCCTCTTTGGACCCCTGTATCTTTAA	324
Qy	1410	CCTCCTGTTTAAAGTTTCTCTTCCAGAAATGAGAGCTGTAAGCTACAGATGGCTTTACA	1469
Db	323	CCTCCTGTTTAAAGTTTGTCTTCCAGAAATCAAAAGTTGTAAGCTACAAATCGTCTTTCA	264
Qy	1470	AATGGAAACCCCA	1481
Db	263	AATGGAACCCCA	252

RESULT 27

RESULI 27
ABN97930

ABN97930
ID ABN97930 standard: DNA: 2784 BP.

XX
ID
ABR 9 / 93 00

AC ABN97930:

XX
XX
XXXXXX

DT 01-AUG-2002 (first entry)

[illegible]

Human retroviral sequence HE

XX
XX

KW Autoimmune d

KW multiple sclerosis; ds.

XXXX

OS Human endogenous retrovirus.

XX

PN WO9967395-A1.

XX

PD 29-DEC-1999.

XX
DE 22 TTY 1000 - 0000 FROA1E13

PF 23-JUN-1999; 99WO-FR001513.

XX
DD 22-JUN-1968. 0000-00007820

PR 23-JUN-1998; 98FR-00007920.
yy

XX
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE

PA (INRM) / INSERM INST NAT SANTE & RECH MEDICALE.
YY

PT Alliel PM. Perin J. Rieger F:

FI
ATTIET FM, FELLIN O, KIEGER F;
XX

DR WPI: 2000-160587/14.

XX
DY 2000-100387/14.
WEI,

PT New nucleic acid sequences of human endogenous retrovirus, HERV-7a, used

PT for diagnosis, treatment and prevention of autoimmune and neurological

PT diseases.

1. **Introduction**

PS Claim 3; Page 133; 225pp; French.

XX	SQ	Sequence	2784 BP; 816 A; 755 C; 535 G; 678 T; 0 U; 0 Other;
		Query Match	83.7%; Score 1239; DB 3; Length 2784;
		Best Local Similarity	91.0%; Pred. No. 0;
		Matches 1364; Conservative	0; Mismatches 100; Indels 35; Gaps 3;
Qy	2	TGGCCCTCCCTTATCATACATCTTTCTCTTTACTGTTCTTTACGCCCTTTTCGCTCTCACTG	61
Db	903	TGGCCCTCCCTTATCATATATTTTCTCTTTACTGTTCTTTACGCCCTTTTCACCTCTCACTG	962
Qy	62	CACCCCTCCATGCTGCTGAC-----AACCAGTAGCTCCCTTAC	102
Db	963	CACCCCTCCATGCGCACTGCACCCCGTCCATGCCCCGTCCATGCGAGTAGCTCCCTTTAG	1022
Qy	103	CAAGAGTTTCTATGAAGAACCGCGCTTCTCTGGAAATATTGATGCCCATCATATAGGAGT	162
Db	1023	CAAGAGTTTCTATGGAGAAATGACGGCTCCCGGAAATATTGATGCCCATTTGTATAGGAGT	1082
Qy	163	TTATCTAAGGGAAATCTCACTTCATCTGCCACACCCCATATGCCCCGCAACTGCTATAAAC	222
Db	1083	TTATCTAAGGGAAACCCCACTTCACTGCGCCACACCCCATATGCCCAACAACCTGCTATAAAC	1142
Qy	223	TCTGCCACTCTTGGCATGATGAAATACATCTATTATTGGACAGGAAAAATGATTAATCCT	282
Db	1143	TCTGCCACTCTTGGCATGCAATGCAATACTCATTTATTGGACAGGAAAAACGATTAATCCC	1202
Qy	283	AGTTGTCTCGAGGACTTGGAGCCTGCTGTTGGACTTACTTCACCCATACCAAGTATG	342
Db	1203	AGTTGTCTCGAGGACTTGGAA-----GGACTCACTTCACTATACAGTATG	1249
Qy	343	TCTGATGGGGTGGAAATTCAGGTTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCC	402
Db	1250	TCTGATGGGGTGGAGTTCAAGATTCAGGCAACAGAAAAACACATAAAGGAAGTAAATCTCC	1309
Qy	403	CAACTGACCCGGGACATAGACACCCCTAGCCCTACAAAGGACTAGTTCTCTCAAACTA	462
Db	1310	CAACTGACCTGGGTACATAGACACCCCTGGCCCTTACAAAGGACTAGATCTCTCAAACTA	1369
Qy	463	CATGAAACCTCCGTACCCATCTCGCGCTGGTGAGCCTATTTAATACCAACCTCACTCGG	522
Db	1370	CATGAAACCTTCATACCCATATCTGGCTGGTAAAGCCTATTTAATACCAACCTGACTGG	1429
Qy	523	CTCCATGAGTCTCAGCCCAAAACCCCTACTAACTGTTGGATGTCCTCCCTCGCACTTC	582
Db	1430	CTCCATGAGTCTCGGCCAAAACCCCTACTAACTGTTGGATGTCCTCCCTCGCACTTT	1489
Qy	583	AGGCCATACATTTCAATCCCTGTTCTCGAAACAATGGNAACAATTCAGCACAGAAATAAC	642
Db	1490	AGGCCATACATTTCAATCCCTATACCTGGAACAATGGNAACAATTCAGCACAGAAATAAC	1549
Qy	643	ACCACTTCCGTTTTAGTAGGACCTCTTGTTTTCAACTCTGGAATAACCCATACCTCAAAC	702
Db	1550	ACCACTTCTGTTTTAGTAGTCTCTCTTTTTCCTGGAATAACCCATACCTCAAAC	1606
Qy	703	CTCACTGTGTAAAAATTTAGCAATACTATAGACACAACCAAGCTCCCAATGCATCAGGTGG	762
Db	1607	CTCACTGTGTAAAAATTTAGCAATACTATAGACACAGCCAACCTCCCAATGCATCAGGTGG	1666
Qy	763	GTAACACCTCCACACGAAATAGTCTGCTTACCTCAGGAAATATTTTTTGTCTGTGTGTAAC	822
Db	1667	GTAACCTCTCCACACGAAATAGTCTGCTTACCTCAGGAAATATTTTTTGTCTGTGTGTAAC	1726
Qy	823	TCAGCCTCATTTGTTGAATGGCTCTTCAGAACTATATGTCCTCTCATTTCTTAGTG	882
Db	1727	TCAGCCTCATTTGTTGAATGGCTCTTCAGAACTCTGTGTCCTCTCATTTCTTAGTG	1786
Qy	883	CCCCATGACCACTACACTGAAACAGATTTTATACAAATCATGTCGTACCTAAGCCCCAC	942
Db	1787	GCCCTATGCCCATCTACATCTGAACAGATTTTATACAACTCATCTACCTAAGCCCCCG	1846
Qy	943	AACAAAAGATGACCAATCTTCTCTTTGTTTATCAGACGAGGAGTGTAGGCAGACTAGGT	1002

Db 765 ACAAAGAGTACCCATCTTCTTCTTTGTTATTGGAGCAGGAGTGCTAGGCGGAGTAGCTA 824
Qy 1004 CTGGCATTGGCAGTATCACAACTCTACTAGTCTTCTACTCAAACTATCTCAAGAAATAA 1063
Db 825 CTGGCATTGGCGGTATCACAACTCTACTAGTCTTCTACTCAAACTATCTCTCAAGAACTAA 884
Qy 1064 ATGTGTACATGGACAGCTCACTGACTCCCTGGTGCACCTGGTGCAGATCAACTTAACCTCC 1123
Db 885 ATGTGTACATGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 944
Qy 1124 TAGCAGCAGTAGTCTCTTCAAAATCGAAGAGCTTTAGACTTTGCTAAACCGCCAAAGAGGGG 1183
Db 945 TAGCAGCAGTAGTCTCTTCAAAATCGAAGAGCTTTAGACTTTGCTAAACCGCGGAAGCGGG 1004
Qy 1184 GAACCTGTTATTTTATAGGAGAGAACGCTGTTATTAATTAATCAATCAAGAAATGTTCA 1243
Db 1005 GAACCTGTTATTTTATAGGAGAGAAATGCTGTTATTAATCAATCAAGAAATGTTCA 1064
Qy 1244 CTGAGAAAGTTAAAGAAATTCGAGATCGAATACAACTAGAGCAGAGGAGCTTCAAAACA 1303
Db 1065 CCGAGAAAGTTAAAGAAATTCAGGTGCAATATTAACGTAGAGCAAGAGGAGCTGCAAAACA 1124
Qy 1304 CCGAACGCTGGGGCTCTCTCAGCCAAATGGATGGCTGGTTCCTCCCTTCTTAGGACCTC 1363
Db 1125 CTGGACCTGGGGCTCTCTCAGCCAAATGGATGGCTGGTTCCTCCCTTCTTAGGACCTC 1184
Qy 1364 TAGCAGCTCTAAATGTTTACTCTCTTTGGACCTCTATCTTTAACTCTCTTTGTTAAGT 1423
Db 1185 TAGCAGCTCTAAATGTTTACTCTCTTTGGACCTCTATCTTTAACTCTCTTTGTTAAGT 1244
Qy 1424 TTCTCTCTCCAGATTTGAAGCTGTAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
Db 1245 TTGTCTTTCCAGAAATCGAAGCAGTAAACCTACAAATCGTCTTCAAAATGGAGCCCCA 1302

RESULT 29

ABAS6337
ID ABAS6337 standard; DNA; 1894 BP.

XX AC ABA56337;

XX 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #4642.

XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000669.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234587P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human foetal liver.

XX Claim 1; SEQ ID NO 4642; 639pp + Sequence Listing; English.

XX

CC The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human foetal liver. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;

Query Match 76.4%; Score 1131.6; DB 4; Length 1894;

Best Local Similarity 92.4%; Pred. No. 0;

Matches 1218; Conservative 0; Mismatches 84; Indels 16; Gaps 2;

Qy 164 TATCTAAGGGAATCTCCACCTTCACCTGCCACACCCATATGCCCGCAACTGCTATTAAT 223

Db 1 TATCTAAGGGAATCTCCACCTTCACCTGCCACACCCATATGCCCGCAACTGCTATTAAT 60

Qy 224 CTGCCACTCTTTGCATGCATGCAAAATACTCATTTATTGGACAGGGAATAATGATTAACTTA 283

Db 61 CTGCCACTCTTTGCATGCATGCAAAATACTCATTTATTGGACAGGGAATAATGATTAACTTA 120

Qy 284 GTTGTCTCTGGAGCACTTGGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAAGTATGT 343

Db 121 GTTGTCTCTGGAGCACTTGGAGCCACTGTCTGTTGGACTTACTTCACCCATACCAAGTATGT 167

Qy 344 CTGATGGGGTGGAAATTCAGAGTTCAGGAGAGAGAAACAAAGTAAAGGAAGCAATCTCCC 403

Db 168 CTGATGGGGTGGAGTTCAGAGTTCAGGAGAGAGAAACAAAGTAAAGGAAGTAACTCTCCC 227

Qy 404 AACTGACCCGGGACATAGCACCCCTAGCCCTTACAAAGGACTAGTCTCTCATAAACTAC 463

Db 228 AACTGACCTGGGTACATAGCACCCCTAGCCCTTACAAAGGACTAGTCTCTCATAAACTAC 287

Qy 464 ATGAAACCCCTCGTACCCATCTCGCTGTGTAGCCCTATTTATACCACTCCTACTCGGC 523

Db 288 ATGAAACCCCTCGTACCCATCTCGCTGTGTAGCCCTATTTATACCACTCCTACTCGGC 347

Qy 524 TCCTAGAGTCTCAGCCCAAAACCTTACTAACTGTTGGATGTGCTCCCTCGGCACCTTCA 583

Db 348 TCCTAGAGTCTCGGCCAAACCTTACTAACTGTTGGATGTGCTCCCTCGGCACCTTCA 407

Qy 584 GGCCATACATTTCAATCCCTGTTCTTGAAACAATGGAACAACCTTCAGCACAGAAATAACA 643

Db 408 GGCCATACATTTCAATCCCTTATACCTGAAACAATGGAACAACCTTCAGCACAGAAATAACA 467

Qy 644 CCACTTCGGTTTGTAGTAGGACCTCTGTTTCCAACTCTGGAATAAACCCATACCTCAAAAC 703

Db 468 CCACTTCGTGTTTGTAGTAGGTCCTC---TTTCCAATCTGGAATAAACCCATACCTCAAAAC 524

Qy 704 TCACCTGTGTAAATTTAGCAATACTATAGACAAACCACTCCCAATGCATCAGGTGGG 763

Db 525 TCACCTGTGTAAATTTAGCAATACTATAGACAAACCACTCCCAATGCATCAGGTGGG 584

Qy 764 TAAACCTCCCAACAGAAATAGTCTGCTACCTCAGCAATATTTTGTCTGTGGTACCT 823

Db 585 TAACTCTCCCAACAGAAATAGTCTGCTACCTCAGCAATATTTTGTCTGTGGTACCT 644

Qy 824 CAGCCTATCATTTGTTGAATGGCTCTTCAAGATCTATGTCTCTCTCATCTTAGTGC 883

Db 645 CAGCCTATCATTTGTTGAATGGCTCTTCAAGATCTGTGTCTCTCTCATCTTAGTGC 704

Qy 884 CCCCTATGACCACTACACTGAAAGATTTATACAACTATGCTGCTACCTCAAGCCCA 943

Db 705 CCCCTATGACCACTACACTGAAAGATTTATACAACTATGCTGCTACCTCAAGCCCA 764

Qy 944 ACAAAAGAGTACCACTCTCTTTTGTGTTATCAGACAGGAGTGTAGGAGAGTAGGTA 1003

Db 765 ACAAAAGAGTACCACTCTCTTTTGTGTTATCAGACAGGAGTGTAGGAGAGTAGGTA 824

Qy 1004 CTGGCAATGGCAGTATCAAAACCTCTACTCAGTCTTACTACAAACTATCTCAAGAAATAA 1063


```
Db      885 ATGCTGACATGGAATGGGTGGCTGATACCTGGTCACCTTGCAGATCAACTTAACTCCC 944
Qy      1124 TAGCAGCAGTAGTCCTTTCAAAATCGAAGAGCTTTAGACTTGTCTAAACGCCCAAAAGAGGGG 1183
Db      945 TAGCAGCAGTAGTCCTTTCAAAATCGAAGAGCTTTAGACTTGTCTAAACGCCGGAAGAGGGG 1004
Qy      1184 GAACCTGTTTATTTTATAGGAGAGAGCGCTGTTATTATGTTAATCAATCCAGATTGTCA 1243
Db      1005 GAACCTTTTATTTTATAGAGGAAAAATGCTGTGTTATGTTAATCAATCCGGAATCATCA 1064
Qy      1244 CTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGAGCTTCAAAACA 1303
Db      1065 CCGAGAAAGTTAAAGAAATTCAGGTGCAATATAACGTAGAGCAAGAGAGCTGCNAACA 1124
Qy      1304 CCGAACGCTGGGGCTCCTCAGCCAATGGATGCCCTGGGTTCTCCCTTCTTAGGACCTC 1363
Db      1125 CTGGACCTGGGGCTCCTCAGCCAATGGATGCCCTGGATTCTCCCTTCTTAGGACCTC 1184
Qy      1364 TAGCAGCTCTAATATTGTTTACTCTCTTTGGACCTGTATCTTTAACCTCCTTGTAAAGT 1423
Db      1185 TAGCAGCTATAATATTGTTTACTCTCTTTGGACCTGTATCTTTAACCTCCTTGTAAAGT 1244
Qy      1424 TTGTCTCTTCCAGAAATGAAGCTGTAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
Db      1245 TTGTCTTTCCAGAAATCGAAGCAGTAATAACTACAAATCGTCTTCAATGGAGCCCA 1302
```

Search completed: January 22, 2005, 16:40:33
Job time : 695.669 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2005, 12:03:14 ; Search time 137.138 Seconds
(without alignments)
7676.074 Million cell updates/sec

Title: US-09-319-156B-9
Perfect score: 1481
Sequence: 1 atggccctcccttatcatcac.....gtcttacaatggaacccca 1481

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1481	100.0	1481	4	US-08-979-847B-105
2	1335.4	90.2	2946	3	Sequence 105, App
3	302	20.4	1329	4	Sequence 3, Appli
4	173.8	11.7	542	1	Sequence 108, App
5	173.8	11.7	542	1	Sequence 48, App
6	145.6	9.8	1722	3	Sequence 2, Appli
7	145.6	9.8	1722	3	Sequence 58, Appl
8	145.6	9.8	1722	4	Sequence 58, Appl
9	144.6	9.8	2908	4	Sequence 54, Appl
10	131.2	8.9	1859	3	Sequence 249, App
11	131.2	8.9	1859	4	Sequence 46, App
12	131.2	8.9	1859	4	Sequence 46, App
13	130.2	8.8	1001	4	Sequence 42, Appl
14	116.8	7.9	1704	1	Sequence 406, App
15	114.8	7.8	80246	3	Sequence 1, Appli
16	114.2	7.7	2518	3	Sequence 4, Appli
17	114.2	7.7	5865	3	Sequence 1, Appli
18	112.2	7.6	4776	4	Sequence 8, Appli
19	106.4	7.2	2809	4	Sequence 1, Appli
20	104.8	7.1	2342	4	Sequence 171, App
21	90.8	6.1	80595	3	Sequence 12, Appl
22	89	6.0	464	3	Sequence 3, Appli
23	89	6.0	3910	3	Sequence 6, Appli
24	70.2	4.7	2002	4	Sequence 1, Appli
25	70	4.7	149	4	Sequence 7, Appli
26	70	4.7	3925	3	Sequence 2428, A
27	68.4	4.6	635	4	Sequence 9, Appli
28					Sequence 102, App

28	68.4	4.6	6028	3	US-09-011-745-5	Sequence 5, Appli
29	68.4	4.6	8060	4	US-08-766-528-1	Sequence 1, Appli
30	68.4	4.6	8060	4	US-09-661-858-1	Sequence 1, Appli
31	68	4.6	10970	3	US-08-716-351A-5	Sequence 5, Appli
32	67	4.5	1911	1	US-08-258-420-8	Sequence 8, Appli
33	67	4.5	1965	1	US-08-258-420-9	Sequence 9, Appli
34	67	4.5	2001	3	US-08-850-961-13	Sequence 13, Appli
35	67	4.5	2001	3	US-09-479-776-13	Sequence 13, Appli
36	67	4.5	2001	4	US-09-315-127-10	Sequence 10, Appli
37	67	4.5	8202	1	US-08-258-420-13	Sequence 13, Appli
38	66.8	4.5	6312	3	US-09-011-745-7	Sequence 7, Appli
39	66.8	4.5	8132	3	US-08-766-528-3	Sequence 3, Appli
40	66.8	4.5	8132	4	US-09-661-858-3	Sequence 3, Appli
41	65.6	4.4	371	4	US-09-821-976-14048	Sequence 14048, A
42	65.4	4.4	6061	3	US-09-011-745-6	Sequence 6, Appli
43	64.2	4.3	1989	3	US-09-232-278A-5	Sequence 5, Appli
44	64.2	4.3	2499	1	US-08-105-483-310	Sequence 310, App
45	64.2	4.3	2499	1	US-08-709-209-310	Sequence 310, App

ALIGNMENTS

RESULT 1

US-08-979-847B-105

; Sequence 105, Application US/08979847B

; Patent No. 6582703

; GENERAL INFORMATION:

; APPLICANT: PERRON, HERVE

; BESEME, FREDERIC

; BEDIN, FREDERIC

; PARANHOS-BACCALA, GLAUCIA

; KOMURIAN-PRADEL, FLORENCE

; JOLIVET-REYNAUD, COLETTE

; MANDRAND, BERNARD

; GABSON, JEREMY

; TUKE, PHILIP

; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL

; THERAPEUTIC PURPOSES

; NUMBER OF SEQUENCES: 210

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OLIFF & BERRIDGE, PLC

; STREET: P.O. BOX 19928

; CITY: ALEXANDRIA

; STATE: VA

; COUNTRY: USA

; ZIP: 22320

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA: US/08/979,847B

; APPLICATION NUMBER: 26-No. 6582703-1997

; FILING DATE: 26-No. 6582703-1997

; CLASSIFICATION: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: BERRIDGE, WILLIAM P.

; REGISTRATION NUMBER: 30,024

; REFERENCE/DOCKET NUMBER: WPB 39046A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-836-6400

; TELEFAX: 703-836-2787

; INFORMATION FOR SEQ ID NO: 105:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1481 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; SEQUENCE DESCRIPTION: SEQ ID NO: 105:

; US-08-979-847B-105

Query Match									
Best Local Similarity 100.0%; Score 1481; DB 4; Length 1481;									
Matches 1481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	ATGGCCCTCCCTTATCATACTTTCTCTTACTGTTCTCTTACCCCTTTGCTCTCACT	60						
DB	1	ATGGCCCTCCCTTATCATACTTTCTCTTACTGTTCTCTTACCCCTTTGCTCTCACT	60						
QY	61	GCACCCCTCCATGCTGTACCAACAGTAGTCCCTTACCAAGAGTTTCTATGAAGA	120						
DB	61	GCACCCCTCCATGCTGTACCAACAGTAGTCCCTTACCAAGAGTTTCTATGAAGA	120						
QY	121	ACCGGCTCTCTGGAATAATGATGCCCCATCATATAGGAGTTTCTTAAGGGAATCC	180						
DB	121	ACCGGCTCTCTGGAATAATGATGCCCCATCATATAGGAGTTTCTTAAGGGAATCC	180						
QY	181	ACCTTCACTGCCACACCCATATGCCCCGCACTGCTATAACTCTGCCACTCTTTGCATG	240						
DB	181	ACCTTCACTGCCACACCCATATGCCCCGCACTGCTATAACTCTGCCACTCTTTGCATG	240						
QY	241	CATGCAATACTCATTAATGCGACGGGAAATGATTAATCTAGTTGCTCTGGAGGACTT	300						
DB	241	CATGCAATACTCATTAATGCGACGGGAAATGATTAATCTAGTTGCTCTGGAGGACTT	300						
QY	301	GGAGCCACTGCTGTTGGACTTACTTCAACCATACCATGATGCTCATGCGGAGTGAAT	360						
DB	301	GGAGCCACTGCTGTTGGACTTACTTCAACCATACCATGATGCTCATGCGGAGTGAAT	360						
QY	361	CAGGTCAGGCAAGAGAAACAAAGTAAAGGAGCAATCTCCCAACTGACCCGGGACAT	420						
DB	361	CAGGTCAGGCAAGAGAAACAAAGTAAAGGAGCAATCTCCCAACTGACCCGGGACAT	420						
QY	421	AGCACCCCTAGCCCTACAAAGGACTAGTTCTCTCAAACTACATGAAACCCCTCGTACC	480						
DB	421	AGCACCCCTAGCCCTACAAAGGACTAGTTCTCTCAAACTACATGAAACCCCTCGTACC	480						
QY	481	CATAGCTGCTGAGGCTGATTTAATACCAACCTCTACTCGGCTCCATGAGTCTCAGCC	540						
DB	481	CATAGCTGCTGAGGCTGATTTAATACCAACCTCTACTCGGCTCCATGAGTCTCAGCC	540						
QY	541	CAAAACCCCTACTAATGTTGATGCTGCTCCCTGCACTTCAGGCGCATACATTTCAATC	600						
DB	541	CAAAACCCCTACTAATGTTGATGCTGCTCCCTGCACTTCAGGCGCATACATTTCAATC	600						
QY	601	CCTGTTCTGAAATGGAACAACTTCAGCAGAAATAAACACCACTTCGTTTTAGTA	660						
DB	601	CCTGTTCTGAAATGGAACAACTTCAGCAGAAATAAACACCACTTCGTTTTAGTA	660						
QY	661	GGACCTCTGTTTCCAAATCTGGAATAACCCATACCTCAAACTCAGCTGTGTAATAATTT	720						
DB	661	GGACCTCTGTTTCCAAATCTGGAATAACCCATACCTCAAACTCAGCTGTGTAATAATTT	720						
QY	721	AGCAATACTATAGACACACAGCTCCCAATGCAATCAGTGGGTAAACCTCCACACGA	780						
DB	721	AGCAATACTATAGACACACAGCTCCCAATGCAATCAGTGGGTAAACCTCCACACGA	780						
QY	781	ATAGTCTGCTACCTCAGGATATTTTGTGCTGTGCTACCTCAGGCTATCATTTGTTG	840						
DB	781	ATAGTCTGCTACCTCAGGATATTTTGTGCTGTGCTACCTCAGGCTATCATTTGTTG	840						
QY	841	AATGGCTCTTCAAGATCTATGCTCTCTCTCAATTTAGTGCCCTCATGACCATCTAC	900						
DB	841	AATGGCTCTTCAAGATCTATGCTCTCTCTCAATTTAGTGCCCTCATGACCATCTAC	900						
QY	901	ACTGAACAGATTTATACATCATGTGTACCTTAAGCCCAACAAAGAGTACCCATT	960						
DB	901	ACTGAACAGATTTATACATCATGTGTACCTTAAGCCCAACAAAGAGTACCCATT	960						
QY	961	CTTCCCTTTGTTATCAGACGAGGTGCTAGGACACTAGGTACTGCGCAITGGCAGTATC	1020						
DB	961	CTTCCCTTTGTTATCAGACGAGGTGCTAGGACACTAGGTACTGCGCAITGGCAGTATC	1020						

QY	1021	ACAACTCTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGGTGACATGGACAG	1080						
DB	1021	ACAACTCTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGGTGACATGGACAG	1080						
QY	1081	GTCACTGACTCCCTGGTCACTTGCAGATCAAACTTAACTCCCTAGCAGCAGTAGTCTCT	1140						
DB	1081	GTCACTGACTCCCTGGTCACTTGCAGATCAAACTTAACTCCCTAGCAGCAGTAGTCTCT	1140						
QY	1141	CAAAATCGAAGAGCTTTAGACTTCTTAACCGCCAAAGAGGGGAACTGTTTATTTT	1200						
DB	1141	CAAAATCGAAGAGCTTTAGACTTCTTAACCGCCAAAGAGGGGAACTGTTTATTTT	1200						
QY	1201	GGAGAAGAACCTGTTATTTATGTTAATCAATCAGAAATGCTCACTAGAAAGTTTAAAGAA	1260						
DB	1201	GGAGAAGAACCTGTTATTTATGTTAATCAATCAGAAATGCTCACTAGAAAGTTTAAAGAA	1260						
QY	1261	ATTGAGATCGAATACATGATAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCTC	1320						
DB	1261	ATTGAGATCGAATACATGATAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCTC	1320						
QY	1321	CTCAGCCAAATGGATGCTTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG	1380						
DB	1321	CTCAGCCAAATGGATGCTTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG	1380						
QY	1381	TTACTCTCTTTGGACCTGATCTTTAACTCTTGTAAAGTTTGTCTCTCCAGAAAT	1440						
DB	1381	TTACTCTCTTTGGACCTGATCTTTAACTCTTGTAAAGTTTGTCTCTCCAGAAAT	1440						
QY	1441	GAAGCTGTAAGCTACAGATGGTCTTACAAATGGAACCCCA	1481						
DB	1441	GAAGCTGTAAGCTACAGATGGTCTTACAAATGGAACCCCA	1481						

RESULT 2

US-09-175-928-3

; Sequence 3, Application US/09175928A

; Patent No. 6312921

; GENERAL INFORMATION:

; APPLICANT: Jacobs, Kenneth

; APPLICANT: McCoy, John M.

; APPLICANT: Lavallie, Edward R.

; APPLICANT: Collins-Racie, Lisa A.

; APPLICANT: Evans, Cheryl

; APPLICANT: Merberg, David

; APPLICANT: Treacy, Maurice

; APPLICANT: Mi, Sha

; APPLICANT: Genetics Institute, Inc.

; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

; FILE REFERENCE: 6006B.AJ172A

; CURRENT APPLICATION NUMBER: US/09/175,928A

; CURRENT FILING DATE: 1998-10-20

; NUMBER OF SEQ ID NOS: 62

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 2946

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-175-928-3

Query Match 90.2%; Score 1335.4; DB 3; Length 2946;

Best Local Similarity 93.9%; Pred. No. 0;

Matches 1390; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1 ATGGCCCTCCCTTATCATACTTTCTTACTGTTCTCTTACCCCTTTGCTCTCACT 60

DB 928 ATGGCCCTCCCTTATCATATTTTCTTACTGTTCTTTTACCCCTTTTCACTCTCACT 987

QY 61 GCACCCCTCCATGCTGTACCAACAGTAGTCCCTTACCAAGAGTTTCTATGAAGA 120

DB 988 GCACCCCTCCATGCTGTACCAACAGTAGTCCCTTACCAAGAGTTTCTATGAAGA 1047

QY 121 ACGCGGCTCTCTGGAATAATGATGCCCCATCATATAGGAGTTTCTAGGAGTATC 180

Db 1048 ATGAGCGTCCGGAAATATTGATGCCCATCGTATAGGAGTCTTCTTAAGGGAACCCCC 1107
Qy 181 ACCTTCACTGCCACACCCCATATGCCCCGCAATGCTGTATTAATCTCTGCACTCTTTGGCATG 240
Db 1108 ACCTTCACTGCCACACCCCATATGCCCCGCAATGCTGTATTAATCTCTGCACTCTTTGGCATG 1167
Qy 241 CATGCAAACTACTATTTATGAGGAGGAAATGATTAATCCTAGTTGCTCTGAGGACTT 300
Db 1168 CATGCAAACTACTATTTATGAGGAGGAAATGATTAATCCTAGTTGCTCTGAGGACTT 1227
Qy 301 GGAGCCACTCTCTGTTGGACTTACTTCCACCACATACAGTATGCTGATGGGGGTGGAATT 360
Db 1228 GGAGTCACTCTGTTGGACTTACTTCCACCACATGCTGATGCTGATGGGGGTGAGTT 1287
Qy 361 CAAGGTCAGGCAAGAGAAAACAAAGTAAGAAAGCAATCTCCCAACTGACCCCGGGACAT 420
Db 1288 CAAGATCAGGCAAGAGAAAACATGTAAGAAAGAAATCTCCCAACTGACCCCGGTACAT 1347
Qy 421 AGCACCCCTAGCCCTACAAAGACTAGTTCTCTCAAACTACATGAACCCCTCCGTACC 480
Db 1348 GGCACTCTAGCCCTACAAAGACTAGATCTCTCAAACTACATGAACCCCTCCGTACC 1407
Qy 481 CATACTCGCTGCTGAGCCCTATTTAATACCAACCCCTCACTCGGCTCCATGAGGTCTCAGCC 540
Db 1408 CATACTCGCTGCTGAGCCCTATTTAATACCAACCCCTCACTCGGCTCCATGAGGTCTCAGCC 1467
Qy 541 CAAAACCTACTAACTGTTGGATGTCCTCCCTGCACTTCCAGGCCATACATTTCAATC 600
Db 1468 CAAAACCTACTAACTGTTGGATGTCCTCCCTGCACTTCCAGGCCATATGTTTCAATC 1527
Qy 601 CTTGTTCTGCAACAAAGTGAACACTTTCAGCACAGAAATAAACCACCTCCGTTTGTAGTA 660
Db 1528 CTTGTTCTGCAACAAAGTGAACACTTTCAGCACAGAAATAAACCACCTCCGTTTGTAGTA 1587
Qy 661 GGACCTCTGTTTCCAACTCGGAAATAAACCACCTCAACCTCACTGTTGTAATAATTT 720
Db 1588 GGACCTCTGTTTCCAACTCGGAAATAAACCACCTCAACCTCACTGTTGTAATAATTT 1647
Qy 721 AGCAATACTATAGACAAACAGCTCCCAATGATCAGGTGGGTAAACACCTCCACACGA 780
Db 1648 AGCAATACTATACAAACCACTCCCAATGATCAGGTGGGTAAACCTCCACACAA 1707
Qy 781 ATAGTCTGCTACCTCAGAAATATTTTGTCTGTGGTACCTCAGGCTATCATTTGTTG 840
Db 1708 ATAGTCTGCTACCTCAGAAATATTTTGTCTGTGGTACCTCAGGCTATCATTTGTTG 1767
Qy 841 AATGGCTCTTCAGAACTATGTCCTCTCTCATTTCTAGTGCCTCCCTATGACCATCTAC 900
Db 1768 AATGGCTCTTCAGAACTATGTCCTCTCTCATTTCTAGTGCCTCCCTATGACCATCTAC 1827
Qy 901 ACTGAACAAGATTTATACATCATGTCTGTAACCTAAGCCCAACAAAGAGTACCCATT 960
Db 1828 ACTGAACAAGATTTATACATTTATGTCATATCTAAGCCCGCAACAAAGAGTACCCATT 1887
Qy 961 CTTCTCTTTGTTATCAGACAGAGTGTCTAGGACAGTACTAGTACTGGCATTTGCCAGTATC 1020
Db 1888 CTTCTCTTTGTTATAGGACAGAGTGTCTAGTGTGCACTAGTACTGGCATTTGCCGTTATC 1947
Qy 1021 ACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGCTGACATGGAACAG 1080
Db 1948 ACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGCGGACATGGAACGG 2007
Qy 1081 GTCACCTGACTCCCTGGTCTACTTTCAGATCAACTTAACTCCCTAGCAGCAGTGTCCCT 1140
Db 2008 GTCGCGGACTCCCTGGTCACTTTCAGATCAACTTAACTCCCTAGCAGCAGTGTCCCT 2067
Qy 1141 CAAAATCGAAGAGCTTTAGACTTGTCTTAAACCGCAAAAGAGGGGGAACCTGTTTATTTTA 1200
Db 2068 CAAAATCGAAGAGCTTTAGACTTGTCTTAAACCGCTGAAAGAGGGGGAACCTGTTTATTTTA 2127
Qy 1201 GGAGAGAGAGCTGTTATTTATGTTATCAATCCAGAAATGCTCACTGAGAAAGTTAAGAA 1260
Db 2128 GGGGAAGAAATGCTGTTATTTATGTTATCAATCCGGAATCGTCACTGAGAAAGTTAAGAA 2187

RESULT 3

US-08-979-847B-108
; Sequence 108, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:

APPLICANT: PERRON, HERVE
; BESEME, FREDERIC
; BEDIN, FREDERIC
; PARANHOS-BACCALA, GLAUCIA
; KOMURIAN-PRADEL, FLORENCE
; JOLIVET-REYNAUD, COLETTE
; MANDRAND, BERNARD
; GARSON, JEREMY
; TURE, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
; THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210

CORRESPONDENCE ADDRESS:

ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-No. 6582703-1997
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 108:

SEQUENCE CHARACTERISTICS:

LENGTH: 1329 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 108:

US-08-979-847B-108

Query Match 20.4%; Score 302; DB 4; Length 1329;
Best Local Similarity 92.7%; Pred. No. 1.1e-87;
Matches 317; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 1140 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCCAAAGAGGGGAAAGCTGTTTATTTT 1199
DB 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCCAAAGAGGGGAAAGCTGTTTATTTT 60
QY 1200 AGGAGAGAAAGCGCTGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1259
DB 61 AGGGGAAGAAAGCTGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 120
QY 1260 AATTCGAGATCGAATCAATATGATAGAGAGAGAGCTTCAAAACACGGAACGCTGGGGCT 1319
DB 121 AATTCGAGATCGAATCAATATGATAGAGAGAGAGCTTCAAAACACGCTGGGGCT 180
QY 1320 CCTCAGCAATGATGCGCTGGGCTTCTCCCTCTTAGAGAGCTCTAGAGCTCTAATAT 1379
DB 181 CCTCAGCAATGATGCGCTGGGCTTCTCCCTCTTAGAGAGCTCTAGAGCTCTAATAT 240
QY 1380 GTTACTCTCTTTGGAGCCCTGATCTTTAACTCTTTGTTAGTTTGTCTCTTCCAGAA 1439
DB 241 TTTACTCTCTTTGGAGCCCTGATCTTTAACTCTTTGTTAGTTTGTCTCTTCCAGAA 300
QY 1440 TGAAGCTGTAAGCTACAGATGCTTTCAAAATGGAACCCCA 1481
DB 301 TGAAGCTGTAAGCTACAAATAGTTCTTCAAAATGGAACCCCA 342

RESULT 4
US-08-686-878A-48
; Sequence 48, Application US/08686878A
; Patent No. 5708157
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Evans, Cheryl
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5951
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 542 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-686-878A-48

Query Match 11.7%; Score 173.8; DB 1; Length 542;

Best Local Similarity 91.5%; Pred. No. 3.5e-46;
Matches 184; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 1 ATGGCCCTCCCTTATCATCTTTCTTTACTGTTCTTTACCCCTTTTCGCTCTCACT 60
DB 342 ATGGCCCTCCCTTATCATCTTTCTTTACTGTTCTTTACCCCTTTTCGCTCTCACT 401
QY 61 GCACCCCTCCATGCTGTGTACCAACCAAGTAGCTCCCTTACCAAGAGTTTCTATGAAGA 120
DB 402 GCACCCCTCCATGCTGTGTGTATGACCAAGTAGCTCCCTTACCAAGAGTTTCTATGAAGA 461
QY 121 AGCGGGCTTCTGGAATATTTGATGCCCCCATCATATAGGAGTTTATCTTAAGGGAACTCC 180
DB 462 ATGACGCGTCCCGGAATATTTGATGCCCCCATCGTATAGGAGTCTTTCTTAAGGGAACTCC 521
QY 181 ACCTTCACTGCCCAACCCAT 201
DB 522 ACCTTCACTGCCCAACCCAT 542

RESULT 5
US-08-721-489-2
; Sequence 2, Application US/08721489
; Patent No. 5786485
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5951
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 542 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-721-489-2

Query Match 11.7%; Score 173.8; DB 1; Length 542;
Best Local Similarity 91.5%; Pred. No. 3.5e-46;
Matches 184; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 1 ATGGCCCTCCCTTATCATCTTTCTTTACTGTTCTTTACCCCTTTTCGCTCTCACT 60
DB 342 ATGGCCCTCCCTTATCATCTTTCTTTACTGTTCTTTACCCCTTTTCGCTCTCACT 401

Qy 61 GCACCCCTCCATGCTGCTGTACAAACAGTAGCTCCCTTACCAGAGATTTCTATGAAGA 120
Db 402 GCACCCCTCCATGCTGCTGTATGACCAAGTAGCTCCCTTACCAGAGATTTCTATGAGA 461
Qy 121 ACGGGGCTTCCTGGAATATTGATGCCCATCATATAGGAGTTTATCTAAGGGAATCTCC 180
Db 462 ATGACGGCTCCGGAATATTGATGCCCATCATATAGGAGTTTATCTAAGGGAATCTCC 521
Qy 181 ACCTTCACTGCCACACCCAT 201
Db 522 ACCTTCACTGCCACACCCAT 542

RESULT 6

US-08-691-563C-58
; Sequence 58, Application US/08691563C
; Patent No. 6001987
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glauca PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/691,563C
; APPLICATION NUMBER: US/08/691,563C
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38588
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1722 base pairs
; TYPE: nucleotide
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-691-563C-58

Query Match 9.8%; Score 145.6; DB 3; Length 1722;
Best Local Similarity 55.7%; Pred. No. 1.1e-36;
Matches 302; Conservative 0; Mismatches 234; Indels 6; Gaps 1;
Qy 940 CACACAAAAGATACCATTCTCTTTGTTATCAGACGAGGTGGCTAGGCACTA 999
Db 346 CAGTGCAGACCCATCAACAAATATCCCTATTTATAGGGTTAGGAATGGCTACTGTCTACA 405
Qy 1000 GGTACTGGCATTTGCAGTATCACAACTCTACTCAGTTCTACTACAACTATCTCAAGAA 1059
Db 406 GGAAGTGAATAGCCGGTTTATCTACTTCTATCTATCTCTACTTACTACCTACTCTCAAGAA 465

Qy 1060 ATAAATGGTGACATGGACAGGTCACTGACTCCCTGGTGCACCTTGCAGATCAACTTAAC 1119
Db 466 TTCTCAGACAGTTTGCAGAAATTAATGAATCTATTCTTACTTTTACAAATCCCAATTAGAC 525
Qy 1120 TCCTTAGCAGCAGTAGTCCCTTCAAAATCGAAGAGCTTTAGACTTGTAAACCGCCAAAAGA 1179
Db 526 TCCTTGGCAGCAATGACTCTCCAAACCGCGAGGGCCACACCTCTCTCACTGCTGAGAAA 585
Qy 1180 GGGGGAACCTGTTTATTTTAGGAGAAGAACGCTGTTTATTTATGTTATCAATCAATCAGAAAT 1239
Db 586 GGAGGACTCTGCACCTTCTTAGGGAAGAGTGTGTTTTTACACTAACCACTCAGGAGATA 645
Qy 1240 GTCACGTAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAA 1299
Db 646 GTACGAGATGCCACCTGGCATTTTACAGGAAGGGCTTCTGATATCAGACAAATGCTTTCA 705
Qy 1300 AACACCGAAGCTGGGGCCCTCTCAGCAATGGATGCCCTGGGTTCTCCCTTCTTAGGA 1359
Db 706 AACTCTTATACCAA-----CCTCTGGAGTTGGCAACATGGCTTCTTCCATTCTAGGT 759
Qy 1360 CTCTAGCAGCTCTAATATTGTTTACTCTCTTTGGACCCCTGTATCTTTAACTCTCTGTT 1419
Db 760 CCATGGCAGCCATCTTGCTGTACTCACCTTTGGGCCCTGTATTTTAAAGCTTCTTGT 819
Qy 1420 AAGTTTGTCTCTCCAGATTGAAGCTGTAAAGCTACAGATGCTTACAAATCGAAGCC 1479
Db 820 AATTTGTTTCTCTAGGATCGAAGCCATCAAGCTACAGATGGTCTTTACAAATGGAACCC 879
Qy 1480 CA 1481
Db 880 CA 881

RESULT 7

US-09-374-766-58
; Sequence 58, Application US/09374766
; Patent No. 6579526
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glauca PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/374,766
; FILING DATE:
; PRIOR APPLICATION DATA: US/08/691,563
; APPLICATION NUMBER: US/08/691,563
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38588
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1722 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-374-766-58

Query Match 9.8%; Score 145.6; DB 4; Length 1722;
Best Local Similarity 55.7%; Pred. No. 1.1e-36;
Matches 302; Conservative 0; Mismatches 234; Indels 6; Gaps 1;

QY 940 CACAAACAAAGAGTACCCATCTCTCTTTGTTATCAGACGAGGAGTGTAGGCAGACTA 999
DB 346 CAGTGCAGAGCCATACAACTAATATCCCTATTTATAGGGTTAGGAATGGCTACTGTACA 405
QY 1000 GGTACTGGCATTCGCGAGTATCAACACCTCTACTCAGTCTTACTACAACTATCTCAAGAA 1059
DB 406 GGAACCTGGAATAGCGGTTTATCTACTTCAATCTTACTTACCATACACTCTCAAGAAT 465
QY 1060 ATAAATGGTGACATGGAAACAGGTCACCTGCTCCCTGGTACCTTGCAGAGTCAACTTAAC 1119
DB 466 TTCTCAGACAGTTTGCAGAAATAATGAAATCTATTCTTACTTTACAAATCCCAATTAGAC 525
QY 1120 TCCCTAGCAGAGTAGTCTTCAAAATCGAAGAGCTTTAGACTGTCTTAACCGCAAAAGA 1179
DB 526 TCTTTGGCAGCAATGACTCTCCAAACCGCGAGGCCACACCTCTCTACTGTGTAGAAA 585
QY 1180 GGGGGAACCTGTTTATTTTAGGAGAAAGCGCTGTATTATGTTATTAATCAATCCAGAAT 1239
DB 586 GGAGGACTGTGCACCTTCTTAGGGAGAGAGTGTGTTTTTACACTAACAGTCAGGATA 645
QY 1240 GTCACTGAGAAAGTTAAAGAAATTCAGATCGAATACAATGTAGCAGAGGAGCTTCAA 1299
DB 646 GTACGAGATGCCACTGGCATTTACAGGAAAGGGCTTCTGATATCAGACAAATGCCTTTCA 705
QY 1300 AACACCGAAGCTGGGGCCCTCTCAGCAATGATGCCCTGGGTCTCCCTCTTCTTAGGA 1359
DB 706 AACTCTTATACCAA-----CCTCTGGAGTTGGCAACATGGCTTCTTCCATTCTAGGT 759
QY 1360 CCTCTAGCAGCTTAATATTTTACTCTCTCTTTGGAACCTGTATCTTTAACTCTCTTGT 1419
DB 760 CCATGCGACCATCTTGTGTTACTCACCTTTGGGCCCTGTATTTTAAAGCTTCTTGTG 819
QY 1420 AAGTTTGTCTTCCAGAAATGAAGCTGTAAGCTACAGATGGTCTTACAAATGGAACCC 1479
DB 820 AAATTTGTTCTCTAGGATCGAAGCCATCAAGCTACAGATGGTCTTACAAATGGAACCC 879

QY 1480 CA 1481
DB 880 CA 881

RESULT 8
US-08-979-847B-54
Sequence 54, Application US/08979847B
Patent No. 6582703
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
BESEME, FREDERIC
BEDIN, FREDERIC
PARANHOS-BACCALA, GLAUCIA
KOMURIAN-PRADEL, FLORENCE
JOLIVET-REYNAUD, COLETTE
MANDRAND, BERNARD
GARSON, JEREMY
TUKU, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210

CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-NO. 6582703-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 1722 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 54:
US-08-979-847B-54

Query Match 9.8%; Score 145.6; DB 4; Length 1722;
Best Local Similarity 55.7%; Pred. No. 1.1e-36;
Matches 302; Conservative 0; Mismatches 234; Indels 6; Gaps 1;

QY 940 CACAAACAAAGAGTACCCATCTCTCTTTGTTATCAGACGAGGAGTGTAGGCAGACTA 999
DB 346 CAGTGCAGAGCCATACAACTAATATCCCTATTTATAGGGTTAGGAATGGCTACTGTACA 405
QY 1000 GGTACTGGCATTCGCGAGTATCAACACCTCTACTCAGTCTTACTACAACTATCTCAAGAA 1059
DB 406 GGAACCTGGAATAGCGGTTTATCTACTTCAATCTTACTTACCATACACTCTCAAGAAT 465
QY 1060 ATAAATGGTGACATGGAAACAGGTCACCTGCTCCCTGGTACCTTGCAGAGTCAACTTAAC 1119
DB 466 TTCTCAGACAGTTTGCAGAAATAATGAAATCTATTCTTACTTTACAAATCCCAATTAGAC 525
QY 1120 TCCCTAGCAGAGTAGTCTTCAAAATCGAAGAGCTTTAGACTGTCTTAAACCGCAAAAGA 1179
DB 526 TCTTTGGCAGCAATGACTCTCCAAACCGCGAGGCCACACCTCTCTACTGTGTAGAAA 585
QY 1180 GGGGGAACCTGTTTATTTTAGGAGAAAGCGCTGTATTATGTTATTAATCAATCCAGAAT 1239
DB 586 GGAGGACTGTGCACCTTCTTAGGGAGAGAGTGTGTTTTTACACTAACAGTCAGGATA 645
QY 1240 GTCACTGAGAAAGTTAAAGAAATTCAGATCGAATACAATGTAGCAGAGGAGCTTCAA 1299
DB 646 GTACGAGATGCCACTGGCATTTACAGGAAAGGGCTTCTGATATCAGACAAATGCCTTTCA 705
QY 1300 AACACCGAAGCTGGGGCCCTCTCAGCAATGATGCCCTGGGTCTCCCTCTTCTTAGGA 1359
DB 706 AACTCTTATACCAA-----CCTCTGGAGTTGGGCAACATGGCTTCTTCCATTCTAGGT 759
QY 1360 CCTCTAGCAGCTTAATATTTTACTCTCTCTTTGGAACCTGTATCTTTAACTCTCTTGT 1419
DB 760 CCATGCGACCATCTTGTGTTACTCACCTTTGGGCCCTGTATTTTAAAGCTTCTTGTG 819
QY 1420 AAGTTTGTCTTCCAGAAATGAAGCTGTAAGCTACAGATGGTCTTACAAATGGAACCC 1479
DB 820 AAATTTGTTCTCTAGGATCGAAGCCATCAAGCTACAGATGGTCTTACAAATGGAACCC 879

QY 1480 CA 1481
Db 880 CA 881

RESULT 9
US-09-799-451-249
; Sequence 249, Application US/09799451
; Patent No. 6783969
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Ma, Yungqing
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Dunrui
; APPLICANT: Yang, Yonghong
; APPLICANT: Wehrman, Tom
; APPLICANT: Ghosh, Reena
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 803
; CURRENT APPLICATION NUMBER: US/09/799,451
; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 948
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 249
; LENGTH: 2908
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (382)..(1995)
US-09-799-451-249

Query Match 9.8%; Score 144.6; DB 4; Length 2908;
Best Local Similarity 56.9%; Pred. No. 3.2e-36;
Matches 286; Conservative 0; Mismatches 214; Indels 3; Gaps 1;

QY 962 TTCCTTTTGTATCAGAGCAGGAGTGCTAGGCAGACTAGTACTGGCATTTGGCAGTATCA 1021
Db 1445 TTCCTTTTGTATCAGAGCAGGAGTGCTAGGCAGACTAGTACTGGCATTTGGCAGTATCA 1504

QY 1022 CAACCTTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGGTGACATGACACAGG 1081
Db 1505 CAAAGCTTCCCTCACCCTATAGCCAGCTCTCAAGGAAATAGCCAAACATTTGACACCA 1564

QY 1082 TCACCTGACTCCCTGGTCACCTTGCAGATCAACTTAACCTCCCTAGCAGCAGTAGTCTTTC 1141
Db 1565 TGGCTAAGCCTTAACGACCATCGAAGCAAAATCGACTCTTTAGCAGCGTAGTCTTTC 1624

QY 1142 AAAATCGAAGCTTTAGACTTGTCTAACCCCAAGAGGGGAACTGTTTATTTTAT 1201
Db 1625 AAAATCGTCCGAGGACTAGACATGTTAACCGCAGCAGCAGGAGGAAATTTGTTGGCCTTAG 1684

QY 1202 GAGAGAACCTGTTATTTATTTATTAATCAATCCAGATTTGTCACCTGAGAAAGTTAAAGAA 1261
Db 1685 ATGAAAAAATGTGCTTTTGGTAAATCAATCAGGAAAGTACAAGCAACATCAGACAC 1744

QY 1262 TTCAGATCGAATACAATGT---AGACGAGGAGCTTTCAAAAACACCGACGCTGGGGCC 1318
Db 1745 TCCTAATCAAGCCTTCAGTTTACGGGAACGAGCCACTAGGGTTGGTTAAATTTGGGAG 1804

QY 1319 TCCTCAGCCAATGGATGCCCCCTGGGTTCTCCCTTCTTTAGGACCTCTAGCAGCTCTAATAT 1378

Db 1805 GAACCTTGGAAATGGTTCTCTTGGGTTCTTCCCTTACAGGCCACCTTGTAGTCTCTAC 1864
QY 1379 TGTACTCTCTCTTTTGGACCCCTGTATCTTTAAACCTCTCTTGTAAAGTTTGTCTCTTCCAGAA 1438
Db 1865 TTTTGTCTCTCTTTTGGTCCATGTCCTCTAATCTAATAACCAATTTGTCTCTCTCGCC 1924

QY 1439 TTCAGCTGTAAAGCTACAGATG 1461
Db 1925 TTCAGGCCATAAAGCTCCAGACG 1947

RESULT 10
US-08-691-563C-46
; Sequence 46, Application US/08691563C
; Patent No. 6001987
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glaucia PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,563C
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38588
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1859 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-691-563C-46

Query Match 8.9%; Score 131.2; DB 3; Length 1859;
Best Local Similarity 57.6%; Pred. No. 5.6e-32;
Matches 257; Conservative 0; Mismatches 183; Indels 6; Gaps 1;

QY 1036 TTCTACTCAAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACCTGACTCCCTG 1095
Db 1092 TACTACACACACTCTCAAGGATTTCTCAGACAGTTTGCAGAAATATGATATCTATC 1151
QY 1096 GTCACCTTGCAGATCAACTTAACCTCCCTAGCAGCAGTAGTCTCTCAAAATCGAAGAGCT 1155
Db 1152 CTTACTCTACATCCCAATAGACTCTTTGGCAGCAGTAGTCTCTCCAAACCGTCAAGGC 1211
QY 1156 TTAGACTTGTAAACCCCAAGAGGGGAACTGTTTATTTTATTTAGAGAAAGACGCTGT 1215

Db	1212	CTAGACCTCTCCTACTGCTGAGAAAGGAGGACTCTGCACCTTCTTAAGGGAAGAGTGTGT	1271
Qy	1216	TATTATGTTAAATCAATCCGAAATGTGACCTGAGAAAGTTAAAGAAATTCGAGATCGAATA	1275
Db	1272	CTTTACACTAACCAAGCTCAGGATAGTATGAGATGCTGCCGGCATTTACAGAAAGGCT	1331
Qy	1276	CAATGTAGACGAGAGGAGCTTCAAAACACCGAACGCTGGGGCTCTCTAGCCCAATGGATG	1335
Db	1332	TCGTGAATCAGACACGCTTTCAAATTC-----CTATACCAACCTCTGGAGTTGGGCA	1385
Qy	1336	CCCTGGTTCCTCCCTCTTAGACCTCTAGAGCTCTAATATTTGTTACTCTCTTTTGA	1395
Db	1386	ACATGGTTTCTCCCTCTCTATGCTCCATGGCTGCCATCTTGCTATTTACTCGGCTTTGG	1445
Qy	1396	CCCTGTATCTTAACTCTCTTGTAGTTTGTCTCTTCCAGAAATGAAGCTGTAAAGCTA	1455
Db	1446	CCCTGTATTTTAACTCTCTTGTCAAAATTTGTTCTTCTTAGGATCGAGGCTCAAGCTA	1505
Qy	1456	CAGATGGTCTTACAAATGGAACCCCA	1481
Db	1506	CAGATGGTCTTACAAATGGAACCCCA	1531

RESULT 11
US-09-374-766-46
; Sequence 46, Application US/09374766
; Patent No. 6579526
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glaucia PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09374,766
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,563
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38589
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1859 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-374-766-46

Query Match	8.9%;	Score 131.2;	DB 4;	Length 1859;
Best Local Similarity	57.6%;	Pred. No. 5.6e-32;		
Matches 257;	Conservative	0;	Mismatches 183;	Indels 6; Gaps 1;
Qy	1036	TTCTACTACAAACTATCTCAAGAAATAAATGGTGTGACATGGAAACAGGTCACTGACTCCCTG	1095	
Db	1092	TACTACCAACACACTCTCAAGAGATTCTCAGACAGTTTGCAGAGAAATAATGATATCTATC	1151	
Qy	1096	GTCACTTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCTCTTCAAAATCGAAGAGCT	1155	
Db	1152	CTTACTCTACAATCCCAAATAGACTCTTTGGCAGCAGTGACTCTTCCAAAACCGTCAAGGC	1211	
Qy	1156	TTAGACTTGTCTAACCGCCAAAAGAGGGGAACTCTTTATTTTATTTAGGAGAAACGCTGT	1215	
Db	1212	CTAGACCTCTCTCACTGCTGAGAAAGGAGGACTCTGCACCTTCTTAAAGGGAAGTGTGT	1271	
Qy	1216	TATTATGTTAAATCAATCCAGAAATGTCAGTGAAGAAAGTTAAAGAAATTCGAGATCGAATA	1275	
Db	1272	CTTTACACTAACCGAGTCAAGGATAGTATGAGATGCTGCCCGGATTTACAGAAAAAGGCT	1331	
Qy	1276	CAATGTAGACGAGAGAGCTTCAAAAACACCGAACGCTGGGGGCTCTCTAGCCCAATGGATG	1335	
Db	1332	TCTGAAATCAGACACGCTTTCAAAATTC-----CTATACCAACCTCTGGAGTTGGGCA	1385	
Qy	1336	CCCTGGGTTCTCCCTCTTTAGGACCTCTAGCAGCTCTAATATTGTTACTCTCTTTTGA	1395	
Db	1386	ACATGGTTCTTCCCTTTCTATGTCCCATGGCTGCGAATCTTGCTATTTACTCGCCTTTGGG	1445	
Qy	1396	CCCTGTATCTTTAACTCTCTTGAAGTTTGTCTCTTCCAGAAATGAAGCTGTAAAGCTA	1455	
Db	1446	CCCTGTATTTTAACTCTCTTGTCAAAATTTGTTCTCTTAGGATCGAGGCCATCAAGCTA	1505	
Qy	1456	CAGATGGTCTTACAAATGGAACCCCA	1481	
Db	1506	CAGATGGTCTTACAAATGGAACCCCA	1531	

RESULT 12
US-08-979-847B-42
; Sequence 42, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GABSON, JEREMY
; APPLICANT: TUBE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,847B
; FILING DATE: 26-No. 6582703-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BERRIDGE, WILLIAM P.

```
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 39046A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1859 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-08-979-847B-42

Query Match      8.8%; Score 131.2; DB 4; Length 1859;
Best Local Similarity 57.6%; Pred. No. 5.6e-32;
Matches 257; Conservative 0; Mismatches 183; Indels 6; Gaps 1;

QY 1036 TTCTACTAGCAACTATCTCAAGAAATAATGGTGACATGGAACAGGTCACTGACTCCCTG 1095
Db 1092 TACTACCACACACTCTCAAGAGGATTTCTCAGACAGTTTGCAGAAATAATGATATCTATC 1151
QY 1096 GTCACTTGCAGATCAACTTAACCTCCTAGCAGCAGTAGTCTTCAAAATCGAAGAGCT 1155
Db 1152 CTTACTCTACAAATCCCAATAGACTCTTTGGCAGCAGTACTCTCCAAACCGTCAAGGC 1211
QY 1156 TTAGACTTGTCAACCGCCAAAGAGGGGGAACCTGTTTATTTTAGGAGAAGAACGCTGT 1215
Db 1212 CTAGACCTCTCACTGCTGAGAAAGGAGGACTCTGCACACTTCTTAAGGGAAGAGTGTGT 1271
QY 1216 TATTATGTTAATCAATCCAGAAATGTCTAGCAAGAGTTTAAAGAAATTCAGATCGAATA 1275
Db 1272 CTTTACACTAACCAAGTCAGGAGTAGTATGATAGTCTGCCCGGCATTTACAGAAAAAGGCT 1331
QY 1276 CAATGTAGACGACGAGGAGCTTCAAAACACCGAGCTGGGGCTCTCAGCCAAATGATG 1335
Db 1332 TCTGAATACAGAACGCTTTCAAAATTC-----CTATACCAACCTCTGGAGTTGGCA 1385
QY 1336 CCCTGGGTTCTCCCTCTTTAGGACCTCTAGCAGCTCTAATATTTGTTACTCTCTTTTGA 1395
Db 1386 ACATGGTTCTTCCCTTCTATGTCCATGCTGCCATCTTGCTATTACTCGCCTTTGG 1445
QY 1396 CCTGTATCTTTAACTCTCTTGTAGTTGTTGCTCTCTCCAGAAATGAAGCTGTAAGCTA 1455
Db 1446 CCCTGTATTTTAACTCTCTTGTCAAAATTTGTTTCTTCTAGGATCGAGGCCATCAAGCTA 1505
QY 1456 CAGATGCTCTTACAAATGGACCCCA 1481
Db 1506 CAGATGCTCTTACAAATGGACCCCA 1531

RESULT 13
US-09-671-317-406
; Sequence 406, Application US/09671317
; Patent No. 6528260
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Cohen, Annick
; FILE OF INVENTION: BIALLELIC MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM
; FILE REFERENCE: 62.US3.CIP
; CURRENT APPLICATION NUMBER: US/09/671,317
; CURRENT FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US 09/536,178
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: PCT/IB00/00403
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/126,269
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/131,961
; PRIOR FILING DATE: 1999-04-30
```

```
; NUMBER OF SEQ ID NOS: 977
; SOFTWARE: Patent.pm
; SEQ ID NO 406
; LENGTH: 1001
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 503
; OTHER INFORMATION: 12-5-158 : polymorphic base C or T
; NAME/KEY: misc binding
; LOCATION: 484..502
; OTHER INFORMATION: 12-5-158.misl
; NAME/KEY: misc binding
; LOCATION: 504..523
; OTHER INFORMATION: 12-5-158.mis2, potential complement
; NAME/KEY: primer bind
; LOCATION: 346..366
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 801..821
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc binding
; LOCATION: 491..515
; OTHER INFORMATION: 12-5-158 potential probe
; NAME/KEY: misc feature
; LOCATION: 314,336,793
; OTHER INFORMATION: n=a, g, c or t
US-09-671-317-406

Query Match      8.8%; Score 130.2; DB 4; Length 1001;
Best Local Similarity 67.8%; Pred. No. 8e-32;
Matches 211; Conservative 0; Mismatches 99; Indels 2; Gaps 2;

QY 773 CCACGCAATAGTCTGCCTACCTCAGG-AAATATTTTGTCTGTGTACTCAGCCTAT 831
Db 688 CCTCAGTTTCACTATCTACCTTCAGGAAATTTTTCATCTGTGATACACAGCCTAT 747
QY 832 CATTTGTTGAATGGCTTTCAGAAATCTATGTGCTTCTCTCATCTCT-TAGTGGCCCCCTAT 890
Db 748 CGATGCCAAAATGGCACTCCAAAAGAACTATGTATGTCTCATTTTCTAGCAGCTCCCAT 807
QY 891 GACCATCTACACTGAACAAGATTATACAATCATGTGCTGCTTAAGCCCCACACAAAG 950
Db 808 GTCCATATATATCTGAACAAGAGTTACAAAGTCTCTTATACCCCAATCTCGCCACACCCG 867
QY 951 AGTACCATTCTTCTCTTTTGTATCAGAGCAGGAGTGTAGGCAGACTAGGTACTGGCAT 1010
Db 868 AGCCCTTATGTGCTTTTACAGTAGGAGGAGGAACTACTAGGCAGGCTTGGGACTGGAT 927
QY 1011 TGSCAGTATCACAACTCTTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGA 1070
Db 928 TGGAGGCATAACCTCTCCACCCTTCTTATAAATTATCATGAGAAATTAATGATGA 987
QY 1071 CATGGAACAGGT 1082
Db 988 CATGGAATGAGT 999

RESULT 14
US-08-007-282B-1
; Sequence 1, Application US/08007282B
; Patent No. 5403582
; GENERAL INFORMATION:
; APPLICANT: NAZERIAN, KEYVAN
; APPLICANT: CALVERT, JAY G.
; APPLICANT: WITTER, RICHARD L.
; APPLICANT: YANAGIDA, NOBORU
; TITLE OF INVENTION: VACCINE COMPRISING FOWLPOX VIRUS
; TITLE OF INVENTION: RECOMBINANTS EXPRESSING THE ENVELOPE GLYCOPROTEIN OF AN
; TITLE OF INVENTION: AVIAN RETICULOENDOTHELIOSIS RETROVIRUS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
```

ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 8110 Gatehouse Road Suite 500 East
CITY: Falls Church
STATE: VA
COUNTRY: USA
ZIP: 22042

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/007,282B
FILING DATE: 19930121

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR., GERALD M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1644-104P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
TELEX: 248345

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO

US-08-007-282B-1
Query Match 7.9%; Score 116.8; DB 1; Length 1704;
Best Local Similarity 51.5%; Pred. No. 2.6e-27;
Matches 268; Conservative 0; Mismatches 252; Indels 0; Gaps 0;

Qy	932	CTAAGCCGCCAACAAGAGTACCCATTCTTCCTTTGTTATCAGAGCAGGAGTGTAG	991
Db	1178	CTAGCGGTCTAAGAGGCGAGTCAGTTATCCCTCTGCTTGGGCTCAGGAGTTTTCAG	1237
Qy	992	GCAGACTAGGTACTGGCATTGGCAGTATCAAACTCTACTCTACTTACTTACAACTAT	1051
Db	1238	GGGTACACTTGTCTGGTGAACGGGGCTTGGGCTCTCGTTTACACTTATCACAAGCTCT	1297
Qy	1052	CTCAAGAAATAAATGGTGACATGGAACAGGTCATGACTCCCTGGTCACTTGCAGATC	1111
Db	1298	CTAATCAATTGATTGAAGATGTCCAGGCTCTTTCCAGGGACCATCAATGACCTACAGACC	1357
Qy	1112	AACTTAATCCCTAGCAGCAGTACTCTTCAAAATCGAAGCTTTAGACTTTGCTAACCG	1171
Db	1358	AGATTGACTCCCTGGCTGAGGTGTCTCGCACTCCAGGAGAGTGTGTTTTTACGTAACAAGT	1417
Qy	1172	CCAAAGAGGGGGAACCTGTTTATTTTAGGAGAAAGCCTGTTATTTATTTATTAATCAAT	1231
Db	1418	CCGACACAGGAGGAATATGTCGCACTCCAGGAGAGTGTGTTTTTACGTAACAAGT	1477
Qy	1232	CCAGAAATGTCACTGAGAAAGTTAAAGAAATTCAGATCGAATACATGTTAGACAGAG	1291
Db	1478	CGGGTATCGTACGTGACAGATTCGAAATCCCAAGAGGACCTTATCGAGAGAAACGTTG	1537
Qy	1292	AGCTTCAAAACACGAGCCTGGGCGCTCTCAGCAATGATGCGCTGGGTTCTCCCT	1351
Db	1538	CACGTGACGAAACCCCTGTGGAGCGCTTGAACGGCTTCTTCCATTTTGTACCT	1597
Qy	1352	TCTTAGGACCTCTAGCAGCTCTAATTAATTTTACTCTCTTTTGGACCCCTGTATCTTAAC	1411
Db	1598	TCTTAGGCCCCCTGTTTGGCTCATATTTTCTTGACCTCGGCGGTGATTATGAAGA	1657
Qy	1412	TCCTGTTTAAGTTTCTCTTCCAGAAATGAAGCTGTAAA	1451
Db	1658	CCCTGACTCGCATTAATACATGACAAATTCAGGCGAGTAAA	1697

RESULT 15

US-09-078-294-4
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 80246
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match 7.8%; Score 114.8; DB 3; Length 80246;
Best Local Similarity 55.7%; Pred. No. 1.3e-25;
Matches 263; Conservative 0; Mismatches 202; Indels 7; Gaps 2;

Qy	1010	TTGGCAGTATCAACCTCTACTCAGTTTCTACTCAAACTATCTCAAGAAATAAATGGTG	1069
Db	58509	TTAACAGCGTAACTATTTATCTTACTACCTCAGACTCTCTCAAGGATTTCTCAGACA	58568
Qy	1070	ACATGGAACAGGTCACTGACTCCTCGTGCACCTTGCAGATCACTTAATCTCCCTAGCAG	1129
Db	58569	GTTTGCAAAAAGAACGAAATCTGCTTACTCTACAATCCCAATAGACTCTTTGGCAG	58628
Qy	1130	CAGTACTCTTCAAAATCGAAGCTTTTAGACTTGTCTAACCCGCAAAAGAGGGAACT	1189
Db	58629	CAGTACTCTCAAAACCCCTGAGGCTAGACTCTCTTACTGCTGAGAAGGAGATTCT	58688
Qy	1190	GTTTATTTTAGGAGAAAGCCTGTTTATTTATTTAATCAATCCAGAAATGTCAGAGA	1249
Db	58689	GCAC-TTCITAGGGGTAGAGTGTGTTTTTATATAACCACTCAGGATAATATGAGATA	58747
Qy	1250	AGTTAAAGAAATTCAGATCGAATACATGATAGCAGCAGGAGCTTCAAAACCCGAAC	1309
Db	58748	CCACCAGGTGTTTACAGGAAAGGCTTCTGAAATCAGACAATGCCTTTTCAAACTCTATA	58807
Qy	1310	GCCTGGGCGCTCTCAGCAATGATGCTCCCTTCTCCCTTCTTTAGGACCTCTAGCAG	1369
Db	58808	CCAA-----CCTCTGGAGTTGGGCGACATGGCTTCTCCCTTCTTAGGTCCTGTGACAG	58861
Qy	1370	CTCTAATATTTTACTCTCTCTTTGGACCTGTATCTTTAACCTCTTGTAAAGTTGTCT	1429
Db	58862	CCATCTTGCTAATAGTCGCAATTTGGGCGCTGATTTTAACTCTTTGGTCAAAATTTGTTT	58921
Qy	1430	CTTCCAGAAATGAAGCTGTAAGCTACAGATGCTTACAAATGGAACCCCA	1481
Db	58922	CCTTAGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGTAAACCCCA	58973

RESULT 16

US-09-011-745-1
; Sequence 1, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; CURRENT FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061

```
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2518
; TYPE: DNA
; ORGANISM: RD114
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)
; OTHER INFORMATION: n is any nucleotide
;
US-09-011-745-1

Query Match          7.7%; Score 114.2; DB 3; Length 2518;
Best Local Similarity 55.0%; Pred. No. 2.3e-26;
Matches 224; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

QY 1045 AAACATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGCTCCCTGGTCACTTG 1104
Db 1834 AAATTATCCCATCAGTTAATATCTGATGTCCTCAAGTCTTATCCGGTACCATACAAGATT 1893
QY 1105 CAAGATCAACTTAACTCCCTAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTTAGACTTG 1164
Db 1894 CAAGACAGGTAGACTGTTAGCTGAAGTAGTTCTCAAAATAGGAGGAGCTGGACCTA 1953
QY 1165 CTAAACCGCCAAAGAGGGGAACCTGTTATTTTATAGGAGAAGAACGCTGTTATATGTT 1224
Db 1954 CTAAACGGCAGAACAGGAGGAATTTGTTAGCTTCAAGAAAATGCTGTTTATGCT 2013
QY 1225 AATCAATCCAGATTTGCTACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGA 1284
Db 2014 AACAACTCAGGAATTTGTGAGAAACAAAATAAGAACCCCTACAAGAAAGAAATTACAAAACGC 2073
QY 1285 GCAGAGGAGCTTCAAAACACCGAAGCTGGGGCTCTCAGCAATGAGTCCCTGGGTT 1344
Db 2074 AGGAAGCTCTGGGACCAACCTCTCTGGACCGGCTGAGGGCTTTCTTCGGTACCTC 2133
QY 1345 CTCCCTCTCTAGGACCTCTAGCAGCTCTAATATTTGTTACTCTCTTTGGACCTGTATC 1404
Db 2134 CTACCTCTCTGGGACCTCTACTCACCCTCTCTACTCATCTAATGAGGATGGGTT 2193
QY 1405 TTAACTCTCTGTTAAGTTTGTCTCTTCCAGAAATGAAGCTGTAA 1451
Db 2194 TTCAGTGGCTCATGGCTTTCATTAATGATAGACTTAATGTTGTACA 2240
```

RESULT 17

```
US-09-011-745-8
; Sequence 8, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; CURRENT FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 5865
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
```

```
; OTHER INFORMATION: construct
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3611)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3612)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3613)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3614)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3799)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3800)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3801)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3802)
; OTHER INFORMATION: n is any nucleotide
;
US-09-011-745-8
```

```
Query Match          7.7%; Score 114.2; DB 3; Length 5865;
Best Local Similarity 55.0%; Pred. No. 3.9e-26;
Matches 224; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

QY 1045 AAACATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGCTCCCTGGTCACTTG 1104
Db 2692 AAATTATCCCATCAGTTAATATCTGATGTCCTCAAGTCTTATCCGGTACCATACAAGATT 2751
QY 1105 CAAGATCAACTTAACTCCCTAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTTAGACTTG 1164
Db 2752 CAAGACAGGTAGACTGTTAGCTGAAGTAGTTCTCAAAATAGGAGGAGCTGGACCTA 2811
QY 1165 CTAAACCGCCAAAGAGGGGAACCTGTTATTTTATAGGAGAAGAACGCTGTTATATGTT 1224
Db 2812 CTAAACGGCAGAACAGGAGGAATTTGTTAGCTTCAAGAAAATGCTGTTTATGCT 2871
QY 1225 AATCAATCCAGAAATTTGCTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGA 1284
Db 2872 ACAAGTCAGGAATTTGTGAGAAACAAAATAAGAACCTTACAAGAAATTTACAAAACGC 2931
QY 1285 GCAGAGGAGCTTCAAAACACCGAAACGCTGGGGCTCTCTCAGCAATGAGATGGCTGGGTT 1344
Db 2932 AGGAAAAGCTTGGCAACCAACCTCTCTGGACCGGGCTGAGGGCTTTCTTCGGTACCTC 2991
QY 1345 CTCCCTCTCTAGGACCTCTAGCAGCTCTAATATTTGTTACTCTCTTTGGACCTGTATC 1404
Db 2992 CTACCTCTCTGGGACCTCTACTCACCCTCTCTACTCATCTAATGAGGATGGGCTGCT 3051
QY 1405 TTAACTCTCTGTTAAGTTTGTCTCTTCCAGAAATGAAGCTGTAA 1451
Db 3052 TTCAGTGGCTCATGGCTTTCATTAATGATAGACTTAATGTTGTACA 3098
```

RESULT 18

```
US-09-555-352-1
; Sequence 1, Application US/09555352
; Patent No. 6544779
; GENERAL INFORMATION:
```

APPLICANT: Cichutek, Klaus
APPLICANT: Mergel-Wilitzer, Heike
TITLE OF INVENTION: PSEUDO-TYPE RETROVIRAL VECTORS WITH
TITLE OF INVENTION: MODIFIABLE SURFACE CAPSID PROTEINS
FILE REFERENCE: 11692-005001
CURRENT APPLICATION NUMBER: US/09/555,352
CURRENT FILING DATE: 2000-08-24
PRIOR FILING DATE: 1998-11-27
PRIOR APPLICATION NUMBER: PCT/DE98/03542
PRIOR FILING DATE: 1997-11-28
PRIOR APPLICATION NUMBER: DE 197 52 855.4
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Fast-Seq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 4776
TYPE: DNA
ORGANISM: Murine leukemia virus
US-09-555-352-1

Query Match 7.6%; Score 112.2; DB 4; Length 4776;
Best Local Similarity 51.5%; Pred. No. 1.5e-25;
Matches 258; Conservative 0; Mismatches 243; Indels 0; Gaps 0;

QY 951 AGTACCCATTCTCTCTTTTGTATATCAGAGCAGGAGTCTAGGAGACTAGGTACTGGCAT 1010
DB 1216 AGTCCAGTTTATCCCTGCTGTGGGTCTAGGGAATTCAGGGGTACACTTCTGCTGGTG 1275
QY 1011 TGGCAGTATCACACCTCTACTAGTCTACTCAACTATCTCAAGAAATAATAGTGA 1070
DB 1276 AACGGGCTTTGGGTCTCCGTTACACTTATCAAGCTCTTAATCAATGATGAAGA 1335
QY 1071 CATGGAACAGTCACTGACTCCCTGGTGCACCTTGCAGATCAACTTAACTCCCTAGCAGC 1130
DB 1336 TGCCAGCTCTTTCAGGACCATCAATGACCTACAGGACCATGACTCCCTGGGTGA 1395
QY 1131 AGTAGTCTTCAAAATCGAAGAGCTTTAGACTTCTTAACCGCAAGAGGGGGAACCTG 1190
DB 1396 GGTGTCTTCAAAATAGAGAGGGTTAGACCTATTGACTGCCGAAACAGGAGGAATATG 1455
QY 1191 TTTATTTTAGGAGAGNACGCTGTATTTATGTTAATCAATCCAGATTTCTCACTGAGAA 1250
DB 1456 TCTCGACTCCAGGAGAGTGTGTTTTTACGCTTAACTGCGGTATCGTACGTGACAA 1515
QY 1251 AGTTAAAGAAATTCGAGATCGAATACAAATGTAGAGCAGAGAGCTTCAAAACACCGAACG 1310
DB 1516 GATCGAAATCTCAAGAGGACCTTATCGAGAGAAACGTCACGTACGACAAACCCCT 1575
QY 1311 CTGGGGCTCTCAGCCATGGATGCCCTGGGTCTCCCTTCTTAGACCTCTAGCAGC 1370
DB 1576 GTGGAGCGCTTGAACCGCTTCTTCCATATTTCTACCTTGTGTAGGCCCTGTGTTGG 1635
QY 1371 TCTAATATTTACTCTCTTTGGACCTGTATCTTTAACTCTTGTAAAGTTTGTCTC 1430
DB 1636 GCTCATATGTTCTTGACCTTCGGCCCGTGCATTTATGAGACCTTGACTGCATATACA 1695
QY 1431 TTCCAGAAATGAAGCTGTAAA 1451
DB 1696 TGACAAATTCAGGAGTAAA 1716

RESULT 19
US-09-620-312D-171
; Sequence 171, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom

APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yuning
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: John Tillinghast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 6569662el Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pt_FL_genes Version 1.0
SEQ ID NO 171
LENGTH: 2809
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1936)..(2364)
US-09-620-312D-171

Query Match 7.2%; Score 106.4; DB 4; Length 2809;
Best Local Similarity 55.0%; Pred. No. 8.6e-24;
Matches 231; Conservative 0; Mismatches 186; Indels 3; Gaps 1;

QY 996 ACTAGTACTGGCATTGGCAGTATCAACACCTCTACTCAGTTCTACTACAACTATCTCA 1055
DB 1851 ACTTGGCACTGGAATAGCAGGCAATAACCACTCAATCACTCATACCAACTATTCAC 1910
QY 1056 AGAATAAATGGTGATGAGACAGTCACTGACTCCCTGGTCACTTGCAGAGTCAACT 1115
DB 1911 AACCCCTTTCTAACCCGTAGAAGATATGCACCTTCCATACCGTCTCCACGACAAAT 1970
QY 1116 TAACTCCCTAGCAGCAGTAGTCTCTTCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCAA 1175
DB 1971 AGACTTCTCTGGTGGAGTCACTCTTCAAACTGGAGAGTCTGGACCTCTTAACCACTGA 2030
QY 1176 AGAGGGGAACTGTATTTATTTTAGGAGAGAAAGCTGTATTTATTTATTTATTTATTT 1235
DB 2031 GAAAGGGGTACCTGTCATATACCTCCAGGAAGAATGCTGTTTCTGTGTTAATGAATCTGG 2090
QY 1236 AATTGCTACTGAGAAAGTTAAAGAAATTCGAGATCG--AATACAATGTAGAGCAGAGA 1292
DB 2091 CATTTGTCATATGCGCAGTTCGTAGGCTTCATGACGGGTGCGAGCTTAGACATCAAGT 2150
QY 1293 GCTTCAAAACACCGAACGCTGGGGCTCTCTCAGCCAAATGGATGCCCTGGTGTCTCCCTT 1352
DB 2151 CGCTGACTCTGTGGCAAGGATCATCCCTCTTAAGATGGATACCTGCTGGTGTGCCCTT 2210
QY 1353 CTTAGGACCTCTAGCAGCTCTAATATTTGTTACTCTCTTTGGACCTCTGTATTTAACTT 1412
DB 2211 CCTAGGACCTCTGATCTTCTCTTCTTCTTACTTAATGATTTGGGCCATGCATATTTAACTT 2270

RESULT 20
US-09-904-615-12
; Sequence 12, Application US/09904615
; Patent No. 6566325
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: P2032P1
; CURRENT APPLICATION NUMBER: US/09/904,615
; CURRENT FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/511,554

; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/097,917
; PRIOR FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: 60/098,634
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 2342
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-904-615-12

Query Match 7.1%; Score 104.8; DB 4; Length 2342;
Best Local Similarity 54.8%; Pred. No. 2.5e-23;
Matches 230; Conservative 0; Mismatches 187; Indels 3; Gaps 1;
QY 996 ACTAGGTACTGGCATTGGCAGTATCAACACTCTACTCAGTCTTCTACTTACAACTATCTCA 1055
DB 1357 ACTTGGCAGTGGATAGCAGGCAATACCACTCAATCACCTCATACCAACACTATTTCAC 1416
QY 1056 AGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTGGTCACTTGCAGATCAACT 1115
DB 1417 AACCTTTTCAACACCGTAGAGATATGCACACTTCCATTACCAGTCTCCACGCAAT 1476
QY 1116 TAACTCCCTAGCAGCAGTACTCTTCAAAATCAAGAGCTTTAGACTTGTCTAAACGCCAA 1175
DB 1477 AGACTTCTCTGGGAGTCACTCTTCAAAAATGGAGAGTCTTGGACCTCTTAACCACTGA 1536
QY 1176 ACAGAGGGAACCTGTTTATTTTGGAGAGAACGCTGTATTATGTTTAATCAATCCAG 1235
DB 1537 GAAGGGGGTACTTGCATATACCTCAGGAAGAATGCTGTTCTGTGTTAATGAATCG 1596
QY 1236 AATTGTCACCTAGAGAAATTAAGAAATTCGAGATCG---AATACAATGTAGACAGAGGA 1292
DB 1597 CATTTGTCATATCGAGTTCGTAGGCTTCATGACAGGCTGCAGAGCTTTGACATCAAGT 1656
QY 1293 GCTTCAAAACACGACGCTGGGGCTCTCAGCCAATGATGCCCTGGGTTCTCCCTT 1352
DB 1657 CGCTGACTCTCTGGTGGCAAGATCATCCCTTCTAAGATGATACCTCTGGTGGTGGCCCTT 1716
QY 1353 CTTAGACCTCTAGCAGCTTAATATTTGTTTACTCTCTTGGACCTGTATCTTAACTT 1412
DB 1717 CTAGGACCCCTGATCTTCTCTCTGTTACTTAATGATTGGGCCATGCAATATTAACTT 1776

RESULT 21
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA
; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

Query Match 6.1%; Score 90.8; DB 3; Length 80595;
Best Local Similarity 57.5%; Pred. No. 8.4e-18;
Matches 206; Conservative 0; Mismatches 142; Indels 10; Gaps 2;
QY 1124 TAGCAGCAGTAGTCTTCAAAATCGAAGCTTTAGACTTGTCTAAACGCCAAAGAGGGG 1183
DB 58859 TGCAGCAGTAGTCTTCAAAACCGCTGAGGCCCTAGACTCTCATGTTGAGAAAGGAAG-- 58916

QY 1184 GAACCTGTTTATTTTATAGGAGAAGACGCTGTTTATTTATGTTTATCAATCCAGAAATTGTCA 1243
DB 58917 --ATTCTGCACCTTTTAGGGGTAGAGTGTGTTTATTTATATAACCAAGTACAGGATAGTAT 58974
QY 1244 CTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAACA 1303
DB 58975 GAGATACCAACCTCTGGAGTTGGCGACATGGCTTCTCCCTTTCTAGGTCTGTG 59034
QY 1304 CCGAACGCTGGGGCTCTCCTCAGCAATGGATGCCCTGGGTTCTCCCTTCTTAGGACCTC 1363
DB 59035 C-----TTATACCAACCTCTGGAGTTGGCGACATGGCTTCTCCCTTTCTAGGTCTGTG 59088
QY 1364 TAGCAGCTCTAATATTGTTTACTCTCTTGGAGCCCTGTATCTTTAACTCTCTTGAAGT 1423
DB 59089 TGACAGCCCATCTTGTCTAATAGTCGCATTTGGGGCCCTGTATTTTAACTCTTGGTCAAT 59148
QY 1424 TTGCTCTTCCAGAAATTTGAAGCTGTAAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
DB 59149 TTGTTTCTCTAGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGTAACCCCA 59206

RESULT 22
US-09-120-653D-6
; Sequence 6, Application US/09120653D
; Patent No. 6385727
; GENERAL INFORMATION:
; APPLICANT: YOON, JI-WON
; APPLICANT: JUN, HEE-SOOK
; APPLICANT: PARK, HAE-JOON
; APPLICANT: AHN, JONG-SEONG
; APPLICANT: HA, YOUNG-JU
; APPLICANT: CHUNG, SOO-IL
; TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
; FILE REFERENCE: 98-338
; CURRENT APPLICATION NUMBER: US/09/120,653D
; CURRENT FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: KR 98-10108
; PRIOR FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 6
; LENGTH: 464
; TYPE: DNA
; ORGANISM: DIABETES-SPECIFIC RETROVIRUS ENDOGENOUS ERV-9
US-09-120-653D-6

Query Match 6.0%; Score 89; DB 3; Length 464;
Best Local Similarity 55.5%; Pred. No. 1.3e-18;
Matches 218; Conservative 0; Mismatches 165; Indels 10; Gaps 2;
QY 1088 ACTCCCTGGTCCACCTTGCAGATCAACTTAACCTCCCTAGCAGCAGTAGTCTTCAAATC 1147
DB 38 ATCTGCTCTTACTTTAGCAATCCAAATAGACACTTTGGCAGCAGTAGTCTCGCAAAACC 97
QY 1148 GAAGAGCTTTAGACTTGTCTAACCGCAAGAGGGGGAACCTGTTTATTTTATAGGAGAAG 1207
DB 98 GCTGAGGCATAGATGTCTCTCACTGCTGAAGAAAGAGGACTCTGCACTTTCTAGGGGAAG 157
QY 1208 AACGCTGTTTATGTTTATCAATCCAGATTTGCTACTGAGAAAGTTAAGAAATTCGAG 1267
DB 158 AGTGTGTTTGTGACACTTAACCGGACGATAGCATGAGAT-----GGCACCAGC 208
QY 1268 ATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGCAACCTGGGGCTCTCTCAGCC 1327
DB 209 GTTTACAGAAAAGGCTGTGAAATGAGTCGCCCTTTTAAATTTTATACCAACCTGTGGT 268
QY 1328 AATGGATGCCCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTGTTACTCC 1387
DB 269 GTGGGCAACATGGCTTCTCACCTTTCTAGGTCCCGTGGCAACCGCTCTGGTGTACTCG 328
QY 1388 TCTTGGACCTGTATCTTTAACTCTCTTGAAGTTGTTGTTCTCTTCCAGAAATGAAGCTG 1447

329 CCTTTGGGCCAGCATTTTTTAAAGTCTTTGTCAAATTGTGTTGTCTAGATTGAGGCCT 387

1448 TAAAGTACAGATGGTCTTTACAAATGGAACCCC 1480

388 TCAAGTACAGATGGTCTTTACATATCAAAACCCC 420

RESULT 23

```

US-09-120-653D-1
; Sequence 1, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
; APPLICANT: YOON, JI-WON
; APPLICANT: JUN, HEE-SOOK
; APPLICANT: PARK, HAE-JOON
; APPLICANT: HA, YONG-SEONG
; APPLICANT: AHN, JONG-JU
; APPLICANT: CHUNG, SOO-IL
; TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
; FILE REFERENCE: 98-338
; CURRENT APPLICATION NUMBER: US/09/120,653D
; CURRENT FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: KR 98-10108
; PRIOR FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 1
; LENGTH: 3910
; TYPE: DNA
; ORGANISM: DIABETES-SPECIFIC ENDOGENOUS RETROVIRUS ERV-9
US-09-120-653D-1

```

Query Match	6.08;	Score 89;	DB 3;	Length 3910;
Best Local Similarity	55.5%;	Pred. No. 4.9e-18;		
Matches 218;	Conservative 0;	Mismatches 165;	Indels 10;	Gaps 2;
QY	1088	ACTCCCTGGTCACTGTGCAAGATCAACTTAACCTCCCTAGCAGCAGTAGTCTCTTCAAATC	1147	
Db	2895	ANCTGTCCTTACTTTTACAATCCAAATAGACACTTTGGCAGCAGTACTCGCAAAACC	2954	
QY	1148	GAAGAGCTTTAGACTTGCTTAACCGCCAAAAGAGAGGGGAACTGTTTATTTTAGGAGAAG	1207	
Db	2955	GCTGAGGCATAGATGTCTCACTGCTGAAGAAAGAGGACTCTGCACCTTTCTTAGGGGAAG	3014	
QY	1208	AACGCTGTTATATGTTTAAATCAATCCAGAAATGTCTAGAGAAAGTTAAAGAAATTCGAG	1267	
Db	3015	AGTGTGTTTGACACTTAACCGGCACGGATAGCATGAGAT-----GGCACCCAGC	3065	
QY	1268	ATCGAATACAATGTAGACGAGAGAGCTTCAAAACACCGACGCTGGGGCTCTCTCAGCC	1327	
Db	3066	GTTTACAGAAAAAGGCTGCTGAATGAGTCGCCCTTTTAAATCTTATACCAACCTGTGGT	3125	
QY	1328	AATGGATGCCCTGGGGTCTCCCTTCTTTAGGACCTCTAGCAGCTCTAATATTGTTACTCC	1387	
Db	3126	GTGGGGCAACATGGCTTCTCACCTTCTAGTCCCGTGGCAACCGCTCTTGTTGTTACTCG	3185	
QY	1388	TCTTTGGACCCCTGATCTTTTAAACCTCCTTGTTAAAGTTTGTCCTCTTCAGAGAAATCGAAGCTG	1447	
Db	3186	CCTTTGGGCCCCAGCAATTTTAAACGTTCTTGTCAAATTTGT-TTGTCFTAGATTTCGAGCGCT	3244	
QY	1448	TAAAGCTACAGATGGTCTTACAAATGGAACCCC	1480	
Db	3245	TCAAGCTACAGATGGTCTTACATCAATCAAAACCCC	3277	

RESULT 24

AS001.124
 US-09-315-127-7
 ; Sequence 7, Application US/09315127
 ; Patent No. 6448390
 ; GENERAL INFORMATION:
 ; APPLICANT: The University of Tennessee, c/o Richard Cox
 ; TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and
 ; TITLE OF INVENTION: Lipoome Vectors and Use in Gene and Drug Therapy
 ; PRIOR APPLICATION NUMBER: US 60/122,487
 ; PRIOR FILING DATE: 1999-02-26
 ; FILE REFERENCE: 59 US2, REG
 ; PATENT No. 6783961
 ; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins

RESULT 25

```

US-09-513-999C-24228
; Sequence 24228, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26

```

```
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 24228
; LENGTH: 149
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 127
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 129
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 135
; OTHER INFORMATION: m=a or c
; US-09-513-999C-24228

Query Match      4.7%; Score 70; DB 4; Length 149;
Best Local Similarity 88.4%; Pred. No. 9.6e-13;
Matches 76; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1396 CCTGTATCTTTAACTCTCTGTTAAAGTTTGTCTCTCCAGAAATGAAGCTGTAAAGCTA 1455
Db 5 CTCTGTATCTTTAACTCTCTGTTAAAGTTTGTCTCTCTCCAGAAATGAAGCTGTAAAGCTA 64

QY 1456 CAGATGCTCTTACAAATGGAAACCCCA 1481
Db 65 CAAATGGTTCTCCAAATGGAGCCCA 90

RESULT 26
US-09-011-745-9
; Sequence 9, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011.745
; CURRENT FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 3925
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; OTHER INFORMATION: construct
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3910)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3911)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3912)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (3913)
; OTHER INFORMATION: n is any nucleotide
; US-09-011-745-9

Query Match      4.7%; Score 70; DB 3; Length 3925;
Best Local Similarity 49.3%; Pred. No. 7.5e-12;
Matches 224; Conservative 0; Mismatches 215; Indels 15; Gaps 1;

QY 1014 CAGTATCACAACTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAAATGGTGACAT 1073
Db 2948 CACTGCCCTAATCAAAACCAGCAGTTTGAGCAGCTTCAGCGCGCTATCCAGACAGACCT 3007

QY 1074 GGNACAGTCACTGACTCCCTGCTGACCTTGCAGATCAACTTAACCTCCCTAGCAGAGT 1133
Db 3008 CAACGAAGTCGAAATAATCAATTAACCAACCTAGAGAAAGTCACTGACCTCGTTGCTGAAGT 3067

QY 1134 AGTCCTTCAAAATCGAAGAGCTTTAGACTTGCTAAACCGCCAAAGAGGGGAAACCTGTTT 1193
Db 3068 AGTCCTACAGAACCGAAGAGGCTAGATTGCTTCTTCTAAAGAGGAGGAGTCTCTGGC 3127

QY 1194 ATTTTGTAGAGAGAAACGCTGTTTATTTATTTAATCAATCCAGAAATGTCTACTGAGAAAGT 1253
Db 3128 AGCCCTAAAGAAAGAAATGTTGTTTTTATGCAGACCAACACGGGACTAGTGAGAGACAGCAT 3187

QY 1254 TAAAGAAATTCGAGATCGAATCAATGTAGAGCAGAGGAGCTTCAAAACACCGACGCTG 1313
Db 3188 GGCCTAAATTAAGGAAAGGCTTAATCAGAGACAAATACTATTTGAGTCAGGCCAAGGTTG 3247

QY 1314 GGGCTCTCT-----CAGCCAAATGGATGCCCTGGGTTCTCCCTCTTCTTAGG 1358
Db 3248 GTTCGAAGGCGAGTTTAATAGATCCCTCGTTTACCACTTAATCTCCACCATCATGGG 3307

QY 1359 ACCTCTAGCAGCTCTAATATTGTTACTCTCTTTTGGACCTGTATCTTTAACTCCCTTGT 1418
Db 3308 ACCTCTAATAGTACTCTTACTGATCTTACTTTTGGACCTGCACTTCTCAATCGATTAGT 3367

QY 1419 TAAGTTTGTCTCTTCCAGAAATGAAGCTGTAAG 1452
Db 3368 TCAATTTGTTAAAGACAGGATCTCAGTAGTCCAG 3401

RESULT 27
US-08-979-847B-102
; Sequence 102, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
; APPLICANT: TURE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
; THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,847B
```

FILING DATE: 26-No. 6582703-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BEBRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 635 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-08-979-847B-102

Query Match 4.6%; Score 68.4; DB 4; Length 635;
Best Local Similarity 87.2%; Pred. No. 7.9e-12;
Matches 75; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1396 CCCTGATCTTTAACTCTCTGTTAGTTTCTCTCCGAGATTGAGCTGTAAGCTA 1455
DB 1 CCCTGATCTTTAACTCTCTGTTAGTTTCTCTCCGAGATTGAGCTGTAAGCTA 60

QY 1456 CAGATGGTCTTACAAATGGACCCCA 1481
DB 61 CAAATGTTCTTCAAATGGAGCACCA 86

RESULT 28
US-09-011-745-5
; Sequence 5, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; CURRENT FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 6028
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; OTHER INFORMATION: construct
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (3774)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (3775)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (3776)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (3777)

OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (3962)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (3963)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (3964)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (3965)
; OTHER INFORMATION: n is any nucleotide
; US-09-011-745-5

Query Match 4.6%; Score 68.4; DB 3; Length 6028;
Best Local Similarity 49.5%; Pred. No. 3.2e-11;
Matches 217; Conservative 0; Mismatches 206; Indels 15; Gaps 1;

QY 1030 ACTCAGTTCTACTACAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGAC 1089
DB 2828 ACCAGCAGTTTGAGCAGCTTCATGCCGTATCCAGACAGACTCAACGAGTCCGAAAG 2887

QY 1090 TCCCTGGTCACCTTGCAAGATCAACTTAACCTCCCTAGCAGCAGTAGTCTTCAAAATCGA 1149
DB 2888 TCAATTACCACCTAGAAAAGTCACCTGACCTCGTTGCTCTGAAGTAGTCTCTACAGAACCGC 2947

QY 1150 AGAGCTTTAGACTTGCTAACCGCAAGAGGGGNAACCTGTTATTTTATAGGAGAGAA 1209
DB 2948 AGAGGCTTAGATTGCTATTCTTAAAGGAGGAGGTCTCTGGCAGCCCTTAAAGAGAA 3007

QY 1210 CGCTGTTATTATGTTAATCAATCCAGAAATGTCACTGAGAAATTAAGAAATTCGAGAT 1269
DB 3008 TGTTGTTTATGACAGCACACAGGGGGCTAGTCAGAGACAGCATGCCAAATTAAGAGAA 3067

QY 1270 CGAATAACAATGTAGACAGAGGAGCTTCAAAACACCGAACGGCTGGGGCTCTCT- 1322
DB 3068 AGCTTAATCAGAGACAAAACCTATTTTGAGACAGGCCAAGGATGGTTTCAAGGGCTGTTT 3127

QY 1323 -----CAGCCAATGGATGCCCTGGGTCTCCCTCTCTTAGGACCTCTAGCAGCTCA 1374
DB 3128 AATAGATCCCCCTGGTTTACCACTTAATCTCCACCATCATGGGACCTCTAATAGTACTC 3187

QY 1375 ATATTGTTACTCTCTTTGGACCCCTGATCTTTAACTCTCTTGTAAAGTTTGTCTCTTC 1434
DB 3188 TTACTGATCTTACTCTTTGGACCTTGACCTTCATTCGATTAGTTCAATTTGTTAAAGAC 3247

QY 1435 AGAATTGAAGCTGTAAG 1452
DB 3248 AGGATCTCAGTAGTCCAG 3265

RESULT 29
US-08-766-528-1
; Sequence 1, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8060 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-766-528-1

Query Match 4.6%; Score 68.4; DB 3; Length 8060;
Best Local Similarity 53.3%; Pred. No. 3.9e-11;
Matches 144; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 1045 AACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCTACTGCTCCCTGGTCACTTG 1104
DB 1523 AACCTACATCGAATTGTTAAACAGAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCTG 1582

QY 1105 CAAGATCAACTTAACCTCCCTAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTTAGACTTG 1164
DB 1583 GAGGAATCCCTAACCTCTTATCTGAAGTAGTCTACAGAAATAGAGAGGTTAGATTTA 1642

QY 1165 CTAACCCCAAAAGAGGGGAACCTGTTATTTTGGAGAGAACGCTGTTATTATGTT 1224
DB 1643 TTATTTCTAAAGAGGAGGATTATGTAGCCCTTGAAGGAGGAATGCTGTTTATGTTG 1702

QY 1225 AATCAATCCAGATTGTCATGAGAAAGTTAAGAAATTCGAGATCGAATACAAATGTAGA 1284
DB 1703 GATCATTGAGGGGCATCAGAGACTCCATGAACAACTTAGAGAAAGGTTGGAGAACGCT 1762

QY 1285 GCAGAGGAGCTTCAAAACACCGAACGCTGG 1314
DB 1763 CGAAGGGAAGGAAGAACTACTCAAGGGTGG 1792

RESULT 30
US-09-661-858-1
Sequence 1, Application US/09661858
Patent No. 6699663
GENERAL INFORMATION:
APPLICANT: Jay A. Fishman
TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
AND METHODS OF USE
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/661,858
FILING DATE: 14-Sep-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/766,528
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8060 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-661-858-1

Query Match 4.6%; Score 68.4; DB 4; Length 8060;
Best Local Similarity 53.3%; Pred. No. 3.9e-11;
Matches 144; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 1045 AACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCTACTGCTCCCTGGTCACTTG 1104
DB 1523 AACCTACATCGAATTGTTAAACAGAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCTG 1582

QY 1105 CAAGATCAACTTAACCTCCCTAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTTAGACTTG 1164
DB 1583 GAGGAATCCCTAACCTCTTATCTGAAGTAGTCTACAGAAATAGAGAGGTTAGATTTA 1642

QY 1165 CTAACCCCAAAAGAGGGGAACCTGTTATTTTGGAGAGAACGCTGTTATTATGTT 1224
DB 1643 TTATTTCTAAAGAGGAGGATTATGTAGCCCTTGAAGGAGGAATGCTGTTTATGTTG 1702

QY 1225 AATCAATCCAGAAATTTGTCATGAGAAAGTTTAAAGAAATTCGAGATCGAATACAAATGTAGA 1284
DB 1703 GATCATTGAGGGGCCATCAGAGACTCCATGAACAACTTAGAGAAAGGTTGGAGAACGCT 1762

QY 1285 GCAGAGGAGCTTCAAAACACCGAACGCTGG 1314
DB 1763 CGAAGGGAAGGAAGAACTACTCAAGGGTGG 1792

Search completed: January 22, 2005, 19:36:59
Job time : 141.138 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2005, 12:12:40 ; Search time 804.34 Seconds
(without alignments)
10579.682 Million cell updates/sec

Title: US-09-319-156b-9
Perfect score: 1481
Sequence: 1 atggccctcccttcatcac.....gtcttcaaaatggaacccca 1481

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
 - 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
 - 3: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
 - 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
 - 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
 - 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
 - 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
 - 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
 - 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
 - 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
 - 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
 - 12: /cgn2_6/ptodata/2/pubpna/US09D_PUBCOMB.seq.*
 - 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
 - 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
 - 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
 - 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
 - 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
 - 18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
 - 19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
 - 20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
 - 21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1481	100.0	1481	8	US-08-979-847-105
2	1481	100.0	1481	15	US-10-114-104-105
3	1481	100.0	2030	16	US-10-637-565-18
4	1389.8	93.8	2074	16	US-10-416-642-4
5	1335.4	90.2	2930	9	US-09-902-535-1
6	1335.4	90.2	2946	13	US-10-114-893-134
7	1335.4	90.2	2946	15	US-10-016-249-3
8	1333.8	90.1	1617	16	US-10-133-036-3
9	1333.8	90.1	8523	10	US-09-854-867-21
10	1333.8	90.1	56093	10	US-09-873-367C-81
11	1324.2	89.4	7582	16	US-10-632-793-30
12	1308.2	88.3	2782	16	US-10-133-036-1

13	1306.6	88.2	2782	16	US-10-632-793-26	Sequence 26, Appl
14	1277.4	86.3	161334	13	US-10-087-192-730	Sequence 730, App
15	1131.6	76.4	1894	9	US-09-864-761-4444	Sequence 4444, Ap
16	1116.4	75.4	1948	16	US-10-632-793-24	Sequence 24, Appl
17	716.6	48.4	792	9	US-09-864-761-21192	Sequence 21192, A
18	591.8	40.0	1684	16	US-10-363-616-228	Sequence 228, App
19	582.8	39.4	591	16	US-10-632-793-19	Sequence 19, Appl
20	516.6	34.9	1136	16	US-10-632-793-25	Sequence 25, Appl
21	513.6	34.7	2006	16	US-10-632-793-23	Sequence 23, Appl
22	377.8	25.5	570	13	US-10-027-632-322491	Sequence 322491,
23	377.8	25.5	570	13	US-10-027-632-322492	Sequence 322492,
24	377.8	25.5	570	15	US-10-027-632-322491	Sequence 322491,
25	377.8	25.5	570	15	US-10-027-632-322492	Sequence 322492,
26	364	24.6	521	15	US-10-029-386-4312	Sequence 4312, Ap
27	353	23.8	551	13	US-10-027-632-322574	Sequence 322574,
28	353	23.8	551	13	US-10-027-632-322575	Sequence 322575,
29	353	23.8	551	15	US-10-027-632-322574	Sequence 322574,
30	353	23.8	551	15	US-10-027-632-322575	Sequence 322575,
31	318.6	21.5	822900	15	US-10-292-798-1393	Sequence 1393, Ap
32	302	20.4	1329	8	US-08-979-847-108	Sequence 108, App
33	302	20.4	1329	15	US-10-114-104-108	Sequence 14030, A
34	296	20.0	426	9	US-09-864-761-14030	Sequence 316412,
35	229.8	15.5	625	13	US-10-027-632-316412	Sequence 316412,
36	229.8	15.5	625	15	US-10-027-632-316412	Sequence 86541, A
37	229.6	15.5	619	13	US-10-027-632-86541	Sequence 86541, A
38	229.6	15.5	619	15	US-10-027-632-86541	Sequence 7501, Ap
39	219.4	14.8	559	9	US-09-864-761-7501	Sequence 20259, A
40	213	14.4	494	15	US-10-029-386-20259	Sequence 7310, Ap
41	194.8	13.2	520	9	US-09-864-761-7310	Sequence 150080,
42	189.8	12.8	740	13	US-10-027-632-150080	Sequence 150080,
43	189.8	12.8	740	15	US-10-027-632-150080	Sequence 85218, A
44	173.8	11.7	527	13	US-10-027-632-85218	Sequence 85218, A
45	173.8	11.7	527	15	US-10-027-632-85218	Sequence 85218, A

ALIGNMENTS

RESULT 1

US-08-979-847-105
; Sequence 105, Application US/08979847
; Publication No. US20030039664A1
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
; APPLICANT: TUKE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE, PLC
; STREET: P. O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,847
; FILING DATE: 26-NOV-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

; NAME: BERRIDGE, WILLIAM P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 39046A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ. ID. NO: 105:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1481 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-979-847-105

Query Match      100.0%; Score 1481; DB 8; Length 1481;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ATGCCCTCCCTTATCATACTTCTTCTTACTCTTCTTACCCCTTTTCGGCTCTCACT 60
DB      1  ATGCCCTCCCTTATCAATCTTCTTACTTGTCTCTACCCCTTCGCTCTCACT 60

QY     61  GCACCCCTCCATGCTGCTGTACAAACAGTAGCTCCCTTACCAAGAGTTTCTATGAAGA 120
DB     61  GCACCCCTCCATGCTGCTGTACAAACAGTAGCTCCCTTACCAAGAGTTTCTATGAAGA 120

QY    121  ACGGGCTTCTGGAATATGTATGCCCCCATATAGGAGTTTATCTAAGGGAAACTCC 180
DB    121  ACGGGCTTCTGGAATATGTATGCCCCCATATAGGAGTTTATCTAAGGGAAACTCC 180

QY    181  ACCTTCACTGCCACACCCATATGCCCGCGCAACTGCTATACTCTGCCACTCTTTGCGATG 240
DB    181  ACCTTCACTGCCACACCCATATGCCCGCGCAACTGCTATACTCTGCCACTCTTTGCGATG 240

QY    241  CATGCAATATCTCATTTATTGACAGGGAAATGATTAACTCTAGTGTGCTGGAGGACTT 300
DB    241  CATGCAATATCTCATTTATTGACAGGGAAATGATTAACTCTAGTGTGCTGGAGGACTT 300

QY    301  GGAGCCACTGTCTGTGGACTTACTTCAACCATACCACTATGCTGTATGGGGGTGGAATT 360
DB    301  GGAGCCACTGTCTGTGGACTTACTTCAACCATACCACTATGCTGTATGGGGGTGGAATT 360

QY    361  CAAGGTGAGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCAACTGACCCGGGGACAT 420
DB    361  CAAGGTGAGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCAACTGACCCGGGGACAT 420

QY    421  AGCACCCCTAGCCCTCAAGAGGACTAGTCTCTCAAAACTACATGAAACCTCCGTACCC 480
DB    421  AGCACCCCTAGCCCTCAAGAGGACTAGTCTCTCAAAACTACATGAAACCTCCGTACCC 480

QY    481  CATACTCGCCTGGTGAGCCTATTTAATAACACCTCTCATCGGCTCCATGAGTCTCAGCC 540
DB    481  CATACTCGCCTGGTGAGCCTATTTAATAACACCTCTCATCGGCTCCATGAGTCTCAGCC 540

QY    541  CAAACCCCTACTAACTGTGGATGTGCTCCCTCCCTGCACTTCAGGCGCATACATTTCAAT 600
DB    541  CAAACCCCTACTAACTGTGGATGTGCTCCCTCCCTGCACTTCAGGCGCATACATTTCAAT 600

QY    601  CCTGTTCTCTGAAACAATGGAAACAACCTTCAGCAAGAAATAAAACCACTTCCGTTTAGTA 660
DB    601  CCTGTTCTCTGAAACAATGGAAACAACCTTCAGCAAGAAATAAAACCACTTCCGTTTAGTA 660

QY    661  GGAACCTTTGTTTCCAATCTGGAATACCCATCTCTCAAACTCACCTGCTGTAATAATT 720
DB    661  GGAACCTTTGTTTCCAATCTGGAATACCCATCTCTCAAACTCACCTGCTGTAATAATT 720

QY    721  AGCAATCTATAGACACACAGCTCCCAATGCAATGAGTGGGTGTAACAACCTCCCAACGA 780
DB    721  AGCAATCTATAGACACACAGCTCCCAATGCAATGAGTGGGTGTAACAACCTCCCAACGA 780

QY    781  ATAGTCTGCTTACCCTCAGGAATATTTTGTGTGGTACCTCAGCCTATCATTTGTTTG 840
DB    781  ATAGTCTGCTTACCCTCAGGAATATTTTGTGTGGTACCTCAGCCTATCATTTGTTTG 840

```


;	PRIOR APPLICATION NUMBER:	US/09/890,340
;	PRIOR FILING DATE:	2001-11-27
;	PRIOR APPLICATION NUMBER:	PCT/IB00/00159
;	PRIOR FILING DATE:	2000-02-15
;	PRIOR APPLICATION NUMBER:	EP 99420041.8
;	PRIOR FILING DATE:	1999-02-15
;	NUMBER OF SEQ ID NOS:	20
;	SOFTWARE:	Patentin version 3.0
;	SEQ ID NO 18	
;	LENGTH:	2030
;	TYPE:	DNA
;	ORGANISM:	MSRV-1 retrovirus
;	FEATURE:	
;	NAME/KEY:	CDS
;	LOCATION:	(1)..(1626)
;	US-10-637-565-18	
Query Match 100.0%; Score 1481; DB 16; Length 2030;		
Best Local Similarity 100.0%; Pred. No. 0;		
Matches 1481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	1	ATGCCCTCCCTTATCATACTTTTCTCTTATCTGTTCTTACCCCTTTCCTTACCCCTTTCGGTCTTCACCT 60
Db	1	ATGCCCTCCCTTATCATACTTTTCTCTTATCTGTTCTTACCCCTTTCCTTACCCCTTTCGGTCTTCACCT 60
Qy	61	GCACCCCTCCATCGTCTGTACAAACAGTAGTCTCCCTTTACCAAGAGTTTCTATGAAGA 120
Db	61	GCACCCCTCCATCGTCTGTACAAACAGTAGTCTCCCTTTACCAAGAGTTTCTATGAAGA 120
Qy	121	ACGGGGCTTCTGGAATATTTGATGCCCCCATCATATAGGAGTTTATCTAAGGGAACTCC 180
Db	121	ACGGGGCTTCTGGAATATTTGATGCCCCCATCATATAGGAGTTTATCTAAGGGAACTCC 180
Qy	181	ACCTTCACTGCCACACCCATATGCCCGCAACTGCTATTAACCTGCCACTCTTTTGGCATG 240
Db	181	ACCTTCACTGCCACACCCATATGCCCGCAACTGCTATTAACCTGCCACTCTTTTGGCATG 240
Qy	241	CATGCAATATCTCATTTATTTGACAGGGAAATGATTAACTCTAGTTGTCTCTGGAGGACTT 300
Db	241	CATGCAATATCTCATTTATTTGACAGGGAAATGATTAACTCTAGTTGTCTCTGGAGGACTT 300
Qy	301	GGAGCCACTGTCTGTGGACTTACTTTCAACCATACAGTATGTCTGTATGGGGGTGGAATT 360
Db	301	GGAGCCACTGTCTGTGGACTTACTTTCAACCATACAGTATGTCTGTATGGGGGTGGAATT 360
Qy	361	CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCAACTGACCGGGGACAT 420
Db	361	CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCAACTGACCGGGGACAT 420
Qy	421	AGCACCCCTAGCCCTTACAAAGGACTGTTCTCTCAAACTACATGAAACCTCCGGTACC 480
Db	421	AGCACCCCTAGCCCTTACAAAGGACTGTTCTCTCAAACTACATGAAACCTCCGGTACC 480
Qy	481	CATACTCGGCTGGTGAGCCCTATTTAATACCAACCTCCTCAGTGGCTCCATGAGTCTCAGCC 540
Db	481	CATACTCGGCTGGTGAGCCCTATTTAATACCAACCTCCTCAGTGGCTCCATGAGTCTCAGCC 540
Qy	541	CAAAACCCCTACTAACTGTTGGATGTGCTCCCGCTGCATCTCAGGCCATACATTTCAATC 600
Db	541	CAAAACCCCTACTAACTGTTGGATGTGCTCCCGCTGCATCTCAGGCCATACATTTCAATC 600
Qy	601	CTGTGTCCTGAAACAAATGGAAACAACTTTCAGCACAGAAATAAAACCACTTTCCGTTTTAGTA 660
Db	601	CTGTGTCCTGAAACAAATGGAAACAACTTTCAGCACAGAAATAAAACCACTTTCCGTTTTAGTA 660
Qy	661	GGACCTCTTGTGTTTCCAATCTGGAATAAACCCCTACCTCAAACCTGCTGTAAATTT 720
Db	661	GGACCTCTTGTGTTTCCAATCTGGAATAAACCCCTACCTCAAACCTGCTGTAAATTT 720
Qy	721	AGCAATACTATAGACACACCACTCCCAATGATCAGGTGGGTAAACACTCCCAACCGA 780
Db	721	AGCAATACTATAGACACACCACTCCCAATGATCAGGTGGGTAAACACTCCCAACCGA 780

```
; NAME/KEY: unsure
; LOCATION: 1995
; OTHER INFORMATION: a, t, c, g, or other
US-10-416-642-4

Query Match      93.8%; Score 1389.8; DB 16; Length 2074;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 1424; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 1 ATGGCCCTCCCTTATCATACTTTCTCTTTACTTGTCTCTTACCCCTTTTCGCTCTCACT 60
DB 64 ATGGCCCTCCCTTATGTATTTTCTCTTTACTTGTCTCTTACCCCTTTTCGCTCTCACT 123
QY 61 GCACCCCTCCATGCTGCTGTACACACAGTAGTCTCCCTTTACCAAGATTTCTATGAGA 120
DB 124 GCACCCCTCCATGCTGCTGTACACACAGTAGTCTCCCTTTACCAAGATTTCTATGAGA 183
QY 121 ACCGGCTCTCCGGAATATTTGATGCCCATCATATAGGAGTTTATCTAAGGGAATCTCC 180
DB 184 ATGGGGCTCCAGACATATTTGATGCCCATCTGTATAGGAGTTTATCTAAGGGAATCTCC 243
QY 181 ACCTTCACTGCCACACCCATATGCCCCGCACTGCTATTAATCTTGCACTCTTTGCAATG 240
DB 244 GCCTTCACTGCCACACCCATATGCCCCGCACTGCTATTAATCTTGCACTCTTTGCAATG 303
QY 241 CATGCAATATCTATTTATGGAAGGGAATGATTAATCTAGTTGCTCTGAGGACTT 300
DB 304 CATGCAATATCTATTTATGGAAGGGAATGATTAATCTAGTTGCTCTGAGGACTT 363
QY 301 GGAGCCACTGCTGTTGGACTTCTTACCCTATACCAGTAGTCTGATGGGGTGAATTT 360
DB 364 GGAGCCACTGCTGTTGGACTTCTTACCCTATACCAGTAGTCTGATGGGGTGAATTT 423
QY 361 CAAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
DB 424 CAAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 483
QY 421 AGCACCCCTAGCCCTTCAAGAGCTAGTTCTCTCAAACTAATGAAACCCCTCGTACC 480
DB 484 AGCACCCCTAGCCCTTCAAGAGCTAGTTCTCTCAAACTAATGAAACCCCTCGTACC 543
QY 481 CATACTGCTGCTGAGGCTTATTTATACCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 544 CATACTGCTGCTGAGGCTTATTTATACCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 603
QY 541 CAAACCCCTACTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
DB 604 CAAACCCCTACTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 663
QY 601 CTGTTCTCTGAAACAATGGGAAACAATTCAGACAGAAATAAACAACCACTTCCGTTTGTAGTA 660
DB 664 CTGTTCTCTGAAACAATGGGAAACAATTCAGACAGAAATAAACAACCACTTCCATTTTGTAGTA 723
QY 661 GGACCTCTGTTTCCATCTGGAATAACCCATACCTCAACCTCAACCTCAACCTCAACCTCAACCT 720
DB 724 GGACCTCTGTTTCCATCTGGAATAACCCATACCTCAACCTCAACCTCAACCTCAACCTCAACCT 783
QY 721 AGCAATACTATAGACACACAGCTCCCAATGCTCAGGTGGGTAAACCTCCCAACCAAGA 780
DB 784 AGCAATACTATAGACACACAGCTCCCAATGCTCAGGTGGGTAAACCTCCCAACCAAGA 843
QY 781 ATAGTCTGCTTACCCTCAGGAATATTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 844 ATAGTCTGCTTACCCTCAGGAATATTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 903
QY 841 AATGGCTCTTCAGAACTATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
DB 904 AATGGCTCTTCAGAACTATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 963
QY 901 ACTGAAACAAGATTTATACATATGCTGCTACCTTAAGCCCAACCAACCAACCAACCAACCAAC 960
DB 964 ACTGAAACAAGATTTATACATATGCTGCTACCTTAAGCCCAACCAACCAACCAACCAACCAAC 1023

; APPLICANT: Keith, Jr., James C.
; APPLICANT: McCoy, John M.
; APPLICANT: M1, Sha
; TITLE OF INVENTION: Methods and compositions for diagnosing
; TITLE OF INVENTION: and treating preclampsia and gestational trophoblast
; FILE REFERENCE: GIN-606B4
; CURRENT APPLICATION NUMBER: US/09/902,535
; CURRENT FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 60/216,657
; PRIOR FILING DATE: 2000-07-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FASCTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2930
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (930)...(2546)
US-09-902-535-1

Query Match      90.2%; Score 1335.4; DB 9; Length 2930;
Best Local Similarity 93.9%; Pred. No. 0;
Matches 1390; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1 ATGGCCCTCCCTTATCATACTTTCTCTTTACTTGTCTCTTACCCCTTTTCGCTCTCACT 60
DB 930 ATGGCCCTCCCTTATCATATTTTCTTCTTACTTGTCTCTTACCCCTTTTCGCTCTCACT 989
QY 61 GCACCCCTCCATGCTGTGTACACACAGTAGTCTCCCTTACCAAGAGTTTCTTATGAGA 120
```

Db 990 GCACCCCTCCATGCGCTGTATGACGAGTAGTCCCTTACCAAGAGTTCTATGGAGA 1049
Qy 121 ACCGGCTTCTTGAATATGATGCCCCCATCATATAGGAGTTTCTAAGGGAATCTCC 180
Db 1050 ATGACGGTCCCGGAATATGATGCCCCCATCGTATAGGAGTCTTCTAAGGGAACCCCC 1109
Qy 181 ACCTTCACTGCCACACCCCATATGCCCCGCAACTGCTATAAATCTCTGCCACTCTTTGCATG 240
Db 1110 ACCTTCACTGCCACACCCCATATGCCCCGCAACTGCTATAAATCTCTGCCACTCTTTGCATG 1169
Qy 241 CATGCAAAATCTCATTTATGGACAGGGAATAATGATTAATCTTAGTTGTCTTGAGGACTT 300
Db 1170 CATGCAAAATCTCATTTATGGACAGGGAATAATGATTAATCTTAGTTGTCTTGAGGACTT 1229
Qy 301 GGAGCCACTGTCTGTTGGACTTACTTACCCTACACAGTATGCTCATGCGGGTGCAATT 360
Db 1230 GGAGTCACTGTCTGTTGGACTTACTTACCCTACACAGTATGCTCATGCGGGTGAGTT 1289
Qy 361 CAAGGTCAGGCAAGAGAAAACAAGTAAGGAAGCAATCTCCCAACTGACCCCGGGGACAT 420
Db 1290 CAAGATCAGGCAAGAGAAAACAAGTAAGGAAGTAACTCTCCCACTCACCCGGGTACAT 1349
Qy 421 AGCACCCCTAGCCCTACAAAGGACTAGTTCTCTCAAACTACATGAACCCCTCCGTACC 480
Db 1350 GGCACTCTAGCCCTACAAAGGACTAGATCTCTCAAACTACATGAACCCCTCCGTACC 1409
Qy 481 CATACTGCGCTGTGAGCTATTATATACCACCTCACTCGGCTCCATGAGGTCTCAGCC 540
Db 1410 CATACTGCGCTGTGAGCTATTATATACCACCTCACTCGGCTCCATGAGGTCTCAGCC 1469
Qy 541 CAAACCCCTACTAATCTTGATGTGCTCCCTCCCTGACCTTCAGGCAATCAATTCATC 600
Db 1470 CAAACCCCTACTAATCTTGATGTGCTCCCTCCCTGACCTTCAGGCAATATGTTCAATC 1529
Qy 601 CTGTGCTCCTGAAACAATGGAAACAATTCAGCACAGAAATAAACGACTCCGTTTGTAGTA 660
Db 1530 CTGTGACTCCTGAAACAATGGAAACAATTCAGCACAGAAATAAACGACTCCGTTTGTAGTA 1589
Qy 661 GGACCTCTGTTTCCCAATCGGAATAACCATACCTCAAACTCACTCTGTGTAATTT 720
Db 1590 GGACCTCTGTTTCCCAATCGGAATAACCATACCTCAAACTCACTCTGTGTAATTT 1649
Qy 721 AGCAATACTATAGACAAACAGCTCCCAATGATCATAGGTGGGTAAACCTCCACACGA 780
Db 1650 AGCAATACTATAGACAAACAGCTCCCAATGATCATAGGTGGGTAACTCTCCACACAA 1709
Qy 781 ATAGTCTGCTTACCTCAGGAATATTTTGTCTGTGGTACCTCAGGCTATCATTTGTTG 840
Db 1710 ATAGTCTGCTTACCTCAGGAATATTTTGTCTGTGGTACCTCAGGCTATCATTTGTTG 1769
Qy 841 AATGGCTCTTCAGAACTATGCTGCTCTCTCATCTTAGTGCCCTCATGACCATCTAC 900
Db 1770 AATGGCTCTTCAGAACTATGCTGCTCTCTCATCTTAGTGCCCTCATGACCATCTAC 1829
Qy 901 ACTGAAACAAGATTATACATCATGCTGATCTAAGCCCCCACAACAAAGAGTACCATT 960
Db 1830 ACTGAAACAAGATTATACATCATGCTGATCTAAGCCCCCACAACAAAGAGTACCATT 1889
Qy 961 CTTCTTTGTTGATCAGACAGGAGTGTAGGACAGCTAGTACTGCGCATGGCAGTATC 1020
Db 1890 CTTCTTTGTTGATCAGGAGGAGTGTAGGTCAGTACTGAGTACTGCGCATGGCGGTATC 1949
Qy 1021 ACACCTCTACTAGTTCTACTACAAATCTCAAGAAATAAATGCTGATGCAATGGAACAG 1080
Db 1950 ACACCTCTACTAGTTCTACTACAAATCTCAAGAAATAAATGCTGATGGAACAG 2009
Qy 1081 GTCACTGACTCCCTGGTCACTTGAAGATCAACTTAACTCCCTAGCAGCAGTACTCTT 1140
Db 2010 GTCGCCGACTCCCTGGTCACTTGAAGATCAACTTAACTCCCTAGCAGCAGTACTCTT 2069
Qy 1141 CAAATCGAAGCTTTAGACTTGTACCTGACCCCAAGAGGGGGAACCTGTTATTTTA 1200

Db 2070 CAAATCGAAGAGCTTTAGACTTGTCTAACCGCTGAAGAGGGGAACCTGTTATTTTA 2129
Qy 1201 GGAGAAGAAGCGTGTATTTATTTAATCAATCCAGAAATGCTGAGAAAGTTTAAAGAA 1260
Db 2130 GGGGAAGAATGCTGTATTATTTAATCAATCCGGAATCGTCACTGAGAAAGTTTAAAGAA 2189
Qy 1261 ATTCCAGATCGAATACATATAGCAGAGGAGCTTCAAAACACCAAGCGCTGGGGCCTC 1320
Db 2190 ATTCCAGATCGAATACATATAGCAGAGGAGCTTCAAAACACCAAGCGCTGGGGCCTC 2249
Qy 1321 CTCAGCCCAATGATGCCCTGGGTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
Db 2250 CTCAGCCCAATGATGCCCTGGATTTCTCCCTTCTTAGGACCTCTAGCAGCTATAATTTG 2309
Qy 1381 TTACTCTCTTTTGGACCCCTGTATCTTTAACTCTTGTAAAGTTGTCTTCCAGAAAT 1440
Db 2310 CTACTCTCTTTTGGACCCCTGTATCTTTAACTCTTGTAACTCTTGTCTTCTTCCAGAAATC 2369
Qy 1441 GAAGCTGTAAAGCTACAGATGCTTCTTACAAATGGGAACCCCA 1481
Db 2370 GAAGCTGTAAAGCTACAAATGGAGCCCAAGATGCAGTCCAA 2410

RESULT 6
US-10-114-893-134
; Sequence 134, Application US/10114893
; Publication No. US20020193567A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Carlin-Duckett, McKeough
; APPLICANT: Kelleher, Kerry S.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: GI 6000-10A
; CURRENT APPLICATION NUMBER: US/10/114,893
; CURRENT FILING DATE: 2002-04-02
; EARLIER APPLICATION NUMBER: 09/413,232
; EARLIER FILING DATE: 1999-10-06
; NUMBER OF SEQ ID NOS: 321
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 134
; LENGTH: 2946
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-114-893-134

Query Match 90.2%; Score 1335.4; DB 13; Length 2946;
Best Local Similarity 93.9%; Pred. No. 0;
Matches 1390; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

Qy 1 ATGGCCCTCCCTTATCATATTTTCTTTTACTGTTCTCTTACCTCCCTTTGCTCTCACT 60
Db 928 ATGGCCCTCCCTTATCATATTTTCTTTTACTGTTCTTTTACTGTTCTTACCTCTTCACTCTCACT 987
Qy 61 GCACCCCTCCATGCTGTGTAACAACAGTAGTCCCTTACCAAGAGTTCTATGAGA 120
Db 988 GCACCCCTCCATGCGGCTGTATGACAGTAGTCCCTTACCAAGAGTTCTATGAGA 1047
Qy 121 ACCGGCTTCTGGAATATGATGCCCATATATAGGAGTTTATCTAAGGGAATCTC 180
Db 1048 ATGACGGTCCCGGAATATGATGCCCATATGATGAGTCTTCTAAGGGAACCCCC 1107
Qy 181 ACCTTCACTGCCACACCCCATATGCCCCGCAACTGCTATAAATCTCTGCACTCTTTGATG 240
Db 1108 ACCTTCACTGCCACACCCCATATGCCCCGCAACTGCTATCACTCTGCACTCTTTGATG 1167

```
QY 241 CATGCAAAATCTATTTATGACAGGGAATAATGATTAATCTAGTTGTCTCTGGAGGACTT 300
Db 1168 CATGCAAAATCTATTTATGACAGGGAATAATGATTAATCTAGTTGTCTCTGGAGGACTT 1227
QY 301 GGAGCCACTCTCTGTTGGACTTACTTCACCCATACACAGTATGCTGATGGGGTGAATTT 360
Db 1228 GGAGTCACTCTGTTGGACTTACTTCACCCAACTGGTATGCTGATGGGGTGGAGTT 1287
QY 361 CAAGGTGAGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGACAT 420
Db 1288 CAAGATCAGGCAAGAGAAAAACAATGTAAGAGAGTAACTCTCCCAACTCACCCGGGTACAT 1347
QY 421 AGACCCCTAGCCCTCAAGAGACTAGTTCTCTCAAACTACATGAACCCCTCCGTACC 480
Db 1348 GGCACCTCTAGCCCTCAAGAGACTAGATCTCTCAAACTACATGAACCCCTCCGTACC 1407
QY 481 CATTACTCGCTGTGAGCTTATTTATACCAACCTCACTCGGCTCCATGAGTCTCAGCC 540
Db 1408 CATTACTCGCTGTGAGCTTATTTATACCAACCTCACTCGGCTCCATGAGTCTCAGCC 1467
QY 541 CAAAAACCTACTAACTGTTGGATGTGCTCCCTCGCACTTCAGGCCATACATTTCAATC 600
Db 1468 CAAAAACCTACTAACTGTTGGATGTGCTCCCTCGCACTTCAGGCCATATGTTCAATC 1527
QY 601 CTTGTTCTTGAAACAATGGAAACAATTCAGCACAGAAATAAACACCACTTCCTGTTTGTGTA 660
Db 1528 CTTGTTCTTGAAACAATGGAAACAATTCAGCACAGAAATAAACACCACTTCCTGTTTGTGTA 1587
QY 661 GGACCTCTGTTTCCCAATCGGAATAACCACTCAACCTCAACCTCAACCTGTTGTAATTT 720
Db 1588 GGACCTCTGTTTCCCAATCGGAATAACCACTCAACCTCAACCTGTTGTAATTT 1647
QY 721 AGCAATCTATAGACACAAACAGCTCCCAATGATCAGGTGGGTAAACCACTCCCAACGA 780
Db 1648 AGCAATCTATAGACACAAACAGCTCCCAATGATCAGGTGGGTAACTCTCTCCCAACGA 1707
QY 781 ATAGTCTGCTACCTCAGAAATATTTTGTCTGTTGTTACCTCAGCTATCAATGTTTGT 840
Db 1708 ATAGTCTGCTACCTCAGAAATATTTTGTCTGTTGTTACCTCAGCTATCGTTGTTG 1767
QY 841 AATGGCTCTTCAGAACTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAC 900
Db 1768 AATGGCTCTTCAGAACTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAC 1827
QY 901 ACTGAAACAAGATTTATACAACTATGCTGTAAGCCCAACCAACCAACCAACCAACCAAT 960
Db 1828 ACTGAAACAAGATTTATACAACTATGCTGTAAGCCCAACCAACCAACCAACCAAT 1887
QY 961 CTTCTCTTTTGTATCAGAGCAGGAGTCTAGGAGACTAGGTACTGGCAATTTGGCAGTATC 1020
Db 1888 CTTCTCTTTTGTATAGAGCAGGAGTCTAGGAGACTAGGTACTGGCAATTTGGCAGTATC 1947
QY 1021 AGAACCTCTACTCAGTCTTACTACAACTATCTCAAGAAATAAATGGTGAATGGAACAG 1080
Db 1948 AGAACCTCTACTCAGTCTTACTACAACTATCTCAAGAAATAAATGGGAGCATGGAACGG 2007
QY 1081 GTCACTGACTCCCTGCTGCTGCAAGATCAACTTAACTCCCTAGCAGCAGTACCTTT 1140
Db 2008 GTCCGCGACTCCCTGCTGCTGCAAGATCAACTTAACTCCCTAGCAGCAGTACCTTT 2067
QY 1141 CAAAATCGAAGAGCTTTAGACTTTGTAAACCGCCAAAGAGGGGGAACCTGTTTATTTTAA 1200
Db 2068 CAAAATCGAAGAGCTTTAGACTTTGTAAACCGCTGAAAGAGGGGGAACCTGTTTATTTTAA 2127
QY 1201 GGAAGAAGACCGTGTATTTATGTTTAAATCAATCCAGAAATTTGTCTCAGAGAAAGTTAAAGAA 1260
Db 2128 GGGGAAGAAGTGTGTTTATTTATGTTTAAATCAATCCGGAATCGTCTCAGAGAAAGTTAAAGAA 2187
QY 1261 ATTTCGAGATCGAATACAACTAGCAGAGGAGCTTTCAAAACACCGAAACGCTGGGGCCTC 1320
Db 2188 ATTTCGAGATCGAATACAACTAGCAGAGGAGCTTTCAAAACACCGAAACGCTGGGGCCTC 2247
```

```
QY 1321 CTCAGCCAAATGGATGCGCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
Db 2248 CTCAGCCAAATGGATGCGCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAATATTG 2307
QY 1381 TTACTCTCTTTTGGACCCCTGTATCTTTAACTCTCTTGTAAAGTTTGTCTCTCCAGAAAT 1440
Db 2308 CTACTCTCTTTTGGACCCCTGTATCTTTAACTCTCTTGTAAAGTTTGTCTCTCCAGAAATC 2367
QY 1441 GAAGCTGTAAAGCTACAGATGGTCTTTACAAATGGAACCCCA 1481
Db 2368 GAAGCTGTAAAGCTACAAATGGAGCCCAAGATCAGTCCAA 2408

RESULT 7
US-10-016-249-3
; Sequence 3, Application US/10016249
; Publication No. US20030100053A1
; GENERAL INFORMATION:
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Mi, Sha
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 6006B.AJ172A
; CURRENT APPLICATION NUMBER: US/10/016,249
; PRIOR FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: US/09/175,928
; PRIOR FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2946
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-016-249-3
```

```
Query Match 90.2%; Score 1335.4; DB 15; Length 2946;
Best Local Similarity 93.9%; Pred. No. 0;
Matches 1390; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1 ATGGCCCTCCCTTATCATCTTTCTTTTACTGTTCTCTTACCCCTTTCTGCTCTCACT 60
Db 928 ATGGCCCTCCCTTATCATCTTTCTTTTACTGTTCTTTTACCTCTTTTCACTCTCACT 987
QY 61 GCACCCCTCCATGCTGTGTACCAACAGTAGCTCCCTTTACCAAGAGTTTCTATGAAGA 120
Db 988 GCACCCCTCCATGCTGTGTACCAACAGTAGCTCCCTTTACCAAGAGTTTCTATGAAGA 1047
QY 121 AGCGGCTCTCTGGAATAATTTGATGCCCATCATATAGAGTTTATTAAGGGAACCTCC 180
Db 1048 ATGCAGCTCCCGGAATATTTGATGCCCATCATATAGAGTTTATTAAGGGAACCTCC 1107
QY 181 ACCTTCACCTGCCACACCCATATGCCCGCACTATAAAGTCTGCACTCTTTGATG 240
Db 1108 ACCTTCACCTGCCACACCCATATGCCCGCACTATAAAGTCTGCACTCTTTGATG 1167
QY 241 CATGCAAAATCTATTTATTTGGACAGGGAATAATGATTAATCTAGTTGTCTCTGGAGGACTT 300
Db 1168 CATGCAAAATCTATTTATTTGGACAGGGAATAATGATTAATCTAGTTGTCTCTGGAGGACTT 1227
QY 301 GGAGCCACTCTCTGTTGGACTTACTTCACCCATACACAGTATGCTGATGGGGTGAATTT 360
Db 1228 GGAGTCACTCTCTGTTGGACTTACTTCACCCAACTGGTATGCTGATGGGGTGGAGTT 1287
QY 361 CAAGGTGAGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGACAT 420
Db 1288 CAAGATCAGGCAAGAGAAAAACAATGTAAGAGAGTAACTCTCCCAACTCACCCGGGTACAT 1347
```



```
QY 1141 CAAAATCGAAGAGCTTTAGACTTGTCTTAACCGCCAAAGAGGGGGAACCTGTTTATTTTAA 1200
Db 37019 CAAAATCGAAGAGCTTTAGACTTGTCTTAACCGCTGAAAGAGGGGGAACCTGTTTATTTTAA 37078
QY 1201 GGAGAAGAAGCGCTTTATTTATTTATTAATCAATCCAGAAATGCTCACTGAGAAAGTTAAAGAA 1260
Db 37079 GGGAAGAAGTCTGTTATTTATTTATTAATCAATCCGGAATGCTCACTGAGAAAGTTAAAGAA 37138
QY 1261 ATTCGAGATCGAATACAATGTAGACGAGAGCTTCAAAACACCGAAGCTGGGCGCTC 1320
Db 37139 ATTCGAGATCGAATACAACGTAGACGAGAGAGCTTCGAAACACTGGACCTCGGGCGCTC 37198
QY 1321 CTGAGCAATGGATGCCCTGGGTTCTCCCTTCTTAGGACCTCTAGCAGCTCTAAATATTG 1380
Db 37199 CTCAGCAATGGATGCCCTGGATTCCTCCCTTCTTAGGACCTCTAGCAGCTCTAAATATTG 37258
QY 1381 TTAATCTCTTTGGACCTGTATCTTTAACTCTCTTGTAAAGTTTGTCTCTCCAGAAAT 1440
Db 37259 CTACTCTCTTTGGACCTGTATCTTTAACTCTCTTGTAAAGTTTGTCTCTCCAGAAAT 37318
QY 1441 GAAGCTGTAAAGCTACAGATGGTCTTTACAAATGGAACCCCA 1481
Db 37319 GAAGCTGTAAAGCTACAAATGGAGCCCAAGATGAGTCCAA 37359

RESULT 11
US-10-632-793-30
; Sequence 30, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucaia
; APPLICANT: MALLET, Francois
; APPLICANT: VOISSET, Cecile
; TITLE OF INVENTION: ENDOGENOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
; TITLE OF INVENTION: AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
; FILE REFERENCE: 110048
; CURRENT APPLICATION NUMBER: US/10/632,793
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: US/09/869,927
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: PCT/FR00/00144
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: FR 95/00888
; PRIOR FILING DATE: 1999-01-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 7582
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (198)..(198)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (307)..(307)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (355)..(355)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1309)..(1309)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1331)..(1331)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
```

```
; LOCATION: (2213)..(2213)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2398)..(2398)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3787)..(3787)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4115)..(4115)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4261)..(4261)
; OTHER INFORMATION: n = a or g or c or t/u
US-10-632-793-30

Query Match 89.4%; Score 1324.2; DB 16; Length 7582;
Best Local Similarity 92.2%; Pred. No. 0;
Matches 1365; Conservative 24; Mismatches 92; Indels 0; Gaps 0;

QY 1 ATGGCCCTCCCTTATCATATTTTCTTCTTACTGTTCTTACCCCTTTCGCTCTCACT 60
Db 5581 ATGGSCCTCCCTTATCATATTTTCTTCTTASTGTTSTTTTACCCCTTTCCTCTCACT 5640
QY 61 GCACCCCTCCATGCTGTGTAACAACAGTAGTCCCTTACCAGAGTTTCTATGAAGA 120
Db 5641 GCACCCCTCCATGCTGTGTAACAACAGTAGTCCCTTACCAGAGTTTCTATGAAGA 5700
QY 121 ACGCCCTCTCTGGAAATATTGATGCCCATCATATAGGAGTTTATTAAGGGAAGTCC 180
Db 5701 ATGCAGCTCCCGAAATATTGATGCCCATCATATAGGAGTCTTSTTAAGGGAAGTCC 5760
QY 181 ACCTTCACCTGCCACACCCATATGCCCCGCACTGCTATAAATCTGCCACTCTTGCATG 240
Db 5761 ACCTTCACCTGCCACACCCATATGCCCCGCACTGCTATAAATCTGCCACTCTTGCATG 5820
QY 241 CATGCAAACTACTCATTATTGGACAGGAAATGATTAACTCTAGTTGTCTCTGGAGGACTT 300
Db 5821 CATGCAAACTACTCATTATTGGACAGGAAATGATTAACTCTAGTTGTCTCTGGAGGACTT 5880
QY 301 GGAGCCACTGTCTGTTGGACTTACTTACCCCAATACAGTATGTCTGATGGGGTGGAAAT 360
Db 5881 GGAGTCACTGTCTGTTGGACTTACTTACCCCAATGCTATGCTGATGGGGTGGAGTT 5940
QY 361 CAAGTCAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Db 5941 CAAGATCAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6000
QY 421 AGCACCCTTAGCCCTTACAAAGGACTAGTTCTCTCAAACTACATGAAGCCCTCCGTACC 480
Db 6001 GGCACTCTAGCCCTTACAAAGGACTAGATCTCTCAAACTACATGAAGCCCTCCGTACC 6060
QY 481 CATACTCGCTGTGTGAGCCCTATTATAACACCCCTCACTCGGCTCCATGAGGTCTCAGCC 540
Db 6061 CATACTCGCTGTGTGAGCCCTATTATAACACCCCTCACTCGGCTCCATGAGGTCTCAGCC 6120
QY 541 CAAAACCTTAACTGTTGGATGTGCTCCCTCCCTGACTTTCAGGCCATACATTTCAATC 600
Db 6121 CAAAACCTTAACTGTTGGATGTGCTCCCTCCCTGACTTTCAGGCCATATGTTTCAATC 6180
QY 601 CTTGTTCTTGAACAATGGAGCAACTTTCAGCAGAGAAATAAACACCTTCCGTTTGTAGTA 660
Db 6181 CTTGTACTTGAACAATGGAGCAACTTTCAGCAGAGAAATAAACACCTTCCGTTTGTAGTA 6240
QY 661 GGAACCTCTTGTTCCTCAATCTGGAATAAACCCATACCTCAAACTACCTGTGTGTAATTT 720
Db 6241 GGAACCTCTTGTTCCTCAATCTGGAATAAACCCATACCTCAAACTACCTGTGTGTAATTT 6300
QY 721 AGCAATACTATAGACACAACACCTGCTCCCAATGATCAGGTGGGTAAACACCTCCACACGA 780
```


Db 1783 ACAAACCTCTACTAGTTCTACTCAAACTATCTCAAGAACTAAATGGGGACATGGAAACGG 1842
Qy 1081 GTCACTGACTCCCTGGTGCACCTTGCAGAGTCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1140
Db 1843 GTCCGCGACTCCCTGGTGCACCTTGCAGAGTCAACTTAACTCCCTAGCAGCAGTAGTCCTT 1902
Qy 1141 CAAATCGAAGAGCTTTAGACTTGTCTAACCAGCAAGAGAGGGGAACCTGTTTATTTTA 1200
Db 1903 CGAAATCGAAGAGCTTTAGACTTGTCTAACCAGCAAGAGAGGGGAACCTGTTTATTTTA 1962
Qy 1201 GGAGAAAGACGCTGTATTATGTTAATCAATCCAGAAATGTCTCACTGAGAAAGTTAAAGAA 1260
Db 1963 GGGGAAGATGCTGTATTATGTTAATCAATCCGAATCGTCACTGAGAAAGTTGAGAA 2022
Qy 1261 ATTCCAGATCGAATACAAATGATAGCAGAGAGGCTTCAAAACACCGAACGCTGGGCGCTC 1320
Db 2023 ATTCCAGATCGAATACAAATGATAGCAGAGAGGCTTCAAAACACCGAACGCTGGGCGCTC 2082
Qy 1321 CTAGCCCAATGGATGCCCTGGTTCCTCCCTCTTAGGACCTCTAGCAGCTCTAATATTG 1380
Db 2083 CTAGCCGATGGATGCCCTGGATTCCTCCCTCTTAGGACCTCTAGCAGCTAATATTG 2142
Qy 1381 TTACTCTCTTTCGACCTCTGATCTTTAACTCCTCTGTTAAAGTTTGTCTTCCAGAAAT 1440
Db 2143 CTACTCTCTTTCGACCTCTGATCTTTGACCTCCTGTTAACTTGTCTTCCAGAAAT 2202
Qy 1441 GAAGCTGTAAAGCTACAGATGGTCTTTACAAATGGAACCCCA 1481
Db 2203 GAAGCTGTAAAGCTACAAATGGAGCCCAAGATGCAGTCCAA 2243

RESULT 13

US-10-632-793-26
; Sequence 26, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: MALLET, Francois
; APPLICANT: VOISSET, Cecile
; TITLE OF INVENTION: ENDOGENOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
; TITLE OF INVENTION: AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
; FILE REFERENCE: 110048
; CURRENT APPLICATION NUMBER: US/10/632,793
; PRIOR FILING DATE: 2003-08-04
; CURRENT APPLICATION NUMBER: US/09/869,927
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: PCT/FR00/00144
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: FR 99/00888
; PRIOR FILING DATE: 1999-01-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
; LENGTH: 2782
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-632-793-26

Query Match 88.2%; Score 1306.6; DB 16; Length 2782;
Best Local Similarity 92.6%; Pred. No. 0;
Matches 1372; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
Qy 1 ATGGCCCTCCCTTATCATCTTTCTTTCTTTACTTGTCTTTACCCCTTTTCGCTCTCACT 60
Db 763 ATGGCCCTCCCTTATCATATTTTCTCTGTAGTGTCTTTACCCCTTTTCACTCTCACT 822
Qy 61 GCACCCCTCCATGCTGCTGTACAAACAGTAGTCCCTTTACCAAGATTTCTATGAAGA 120
Db 823 GCACCCCTCCATGCGCTGTATGACCAAGTAGTCCCTTCAACAGAGTTTCTATGAGA 882
Qy 121 ACGGGCTTCTCGAAATATTGATGCCCATCATATAGGAGTTTATCTAAGGAAACTCC 180

Db 883 ATGCGGTCCTCCGAAATATTGATGCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCC 942
Qy 181 ACCTTCACTGCCACACCCCATATGCCCCGCAACTGTATATAACTCTGTGCCACTCTTTGATG 240
Db 943 ACCTTCACTGCCACACCCCATATGCCCCGCAACTGTATATAACTCTGTGCCACTCTTTGATG 1002
Qy 241 CATGCAAACTACTCATTTATTGGACAGGAAATCATTAATCCTAGTTGTCTCTGGAGGACTT 300
Db 1003 CATGCAAACTACTCATTTATTGGACAGGAAATCATTAATCCTAGTTGTCTCTGGAGGACTT 1062
Qy 301 GGAGCCACTGTCTGTGGGACTTACTTCAACCATAACAGTATGTCTGATGGGGGTGGAATT 360
Db 1063 GGAGTCACTGTCTGTGGGACTTACTTCAACCATAACAGTATGTCTGATGGGGGTGGAATT 1122
Qy 361 CAAGGTGAGCAAGAGAAAAAACAAGTAAAGAAAGAAATCTCCCAACTGACCCGGGGGACAT 420
Db 1123 CAAGATCAGCAGAGAGAAAAACAATGTAAAGAAAGTAAATCTCCCAACTGACCCGGGGTACAT 1182
Qy 421 AGCACCCCTAGCCCTTACAAAGGACTAGTTCTCTCAAACTACATGAAACCTCCCGTACC 480
Db 1183 GGACCTCTAGCCCTTACAAAGGACTAGATCTCTCAAACTACATGAAACCTCCCGTACC 1242
Qy 481 CATACTCGCTGTGAGCCCTATTATAACACCCCTCACTCGGCTCATAGAGGTCTCAGCC 540
Db 1243 CATACTCGCTGTGAGCCCTATTATAACACCCCTCACTCGGCTCATAGAGGTCTCAGCC 1302
Qy 541 CAAAACCTACTAATCTGTTGGATGTCTCCCTCGACTTCAGGCGCATACATTTCAATC 600
Db 1303 CAAAACCTACTAATCTGTTGGATGTCTCCCTCGACTTCAGGCGCATATGTTTCAATC 1362
Qy 601 CTGTCTCTGAAACAAATGGAAACAACTTCAGCAGAGAAATAAACACCACTTCCGTTTGTAT 660
Db 1363 CTGTACTCTGAAACAAATGGAAACAACTTCAGCAGAGAAATAAACACCACTTCCGTTTGTAT 1422
Qy 661 GGACCTCTTGTCTTCCAAATCTGGAATAAACCCATACCTCAAACTCACTCTGTGTAAATTT 720
Db 1423 GGACCTCTTGTCTTCCAAATCTGGAATAAACCCATACCTCAAACTCACTCTGTGTAAATTT 1482
Qy 721 AGCAATACTATAGACAACAACCACTCCCAATGCATCAGTGGGTAAACACTCCCAACGA 780
Db 1483 AGCAATACTATAGACAACAACCACTCCCAATGCATCAGTGGGTAAACACTCCCAACGA 1542
Qy 781 ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGTGTTACCTCAGCCTATCATTTGTTG 840
Db 1543 ATAGTCTGCTACCTCAGGAATATTTTGTCTGTGTGTTACCTCAGCCTATCATTTGTTG 1602
Qy 841 AATGGCTCTTTCAGAACTTATGTCTCTCTCATTTCTAGTGCCCTCATGACCATCTAC 900
Db 1603 AATGGCTCTTTCAGAACTTATGTCTCTCTCATTTCTAGTGCCCTCATGACCATCTAC 1662
Qy 901 ACTGAAACAAGATTTATACAATCATGTCTGTAACCTTAAGCCCCCAACAAAGAGTAGTACCATT 960
Db 1663 ACTGAAACAAGATTTATACAGTTATGTCTATCTAAGCCCCCAACAAAGAGTAGTACCATT 1722
Qy 961 CTTCTCTTTTGTATCAGCAGAGAGTGTAGGAGACTAGTACTAGCATTGGCAGTATC 1020
Db 1723 CTTCTCTTTTGTATCAGCAGAGAGTGTAGGAGTGTAGTACTAGTACTGGCATTGGCGGTATC 1782
Qy 1021 ACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGTTGATGGAACAG 1080
Db 1783 ACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGTTGGAACAG 1842
Qy 1081 GTCACTGACTCCCTGGTGCACCTTTCGAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCCTT 1140
Db 1843 GTCCGCGACTCCCTGGTGCACCTTTCGAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCCTT 1902
Qy 1141 CAAATCGAAGAGCTTTAGACTTGTCTAACCAGCAAGAGAGGGGAACCTGTTTATTTTA 1200
Db 1903 CGAAATCGAAGAGCTTTAGACTTGTCTAACCAGCAAGAGAGGGGAACCTGTTTATTTTA 1962
Qy 1201 GGNAGAGAGCGCTGTTTATTTATGTTAATCAATCCAGAAATGTCTCACTGAGAAAGTTAAGAA 1260
Db 1963 GGGGAAGAAATGCTGTATTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAGAA 2022

QY 1261 ATTCCGATCGAATACAACTAGACGAGAGAGCTTCAAAACACCGAAACGCTGGGGCCTC 1320
DB 2023 ATTCCGATCGAATACAACTAGACGAGAGAGCTTCAAAACACCGAAACGCTGGGGCCTC 2082
QY 1321 CTCAGGCAATGAGTGGCTGGTTCCTCCCTCTTCTAGGACCTCTAGCAGCTCTAATATTG 1380
DB 2083 CTCAGGCGATGAGTGGCTGGTTCCTCCCTCTTCTAGGACCTCTAGCAGCTCTAATATTG 2142
QY 1381 TTAATCTCTTTGGACCTGATCTTTAACTCTCTTGAAGTTTCTCTTCCAGAAAT 1440
DB 2143 CTACTCTCTTTGGACCTGATCTTTGAACCTCTCTTGAAGTTTCTCTTCCAGAAAT 2202
QY 1441 GAAGCTGTAAAGCTACAGATGGTCTTCAAAATGGAACCCCA 1481
DB 2203 GAAGCTGTAAAGCTACAGATGGTCTTCAAAATGGAACCCCA 2243

RESULT 14
US-10-087-192-730/c
; Sequence 730, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 730
; LENGTH: 161334
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(161334)
; OTHER INFORMATION: n = A,T,C or G

US-10-087-192-730

Query Match 86.3%; Score 1277.4; DB 13; Length 161334;
Best Local Similarity 92.5%; Pred. No. 0;
Matches 1399; Conservative 0; Mismatches 81; Indels 33; Gaps 4;
QY 1 ATGGCCCTCCCTTATCATATTTTCTTTACTGTTCTTTACCCCTTTTGGCTCTCACT 60
DB 102037 ATGGCCCTCCCTTATCATATTTTCTTTACTGTTCTTTACCCCTTTTGGCTCTCACT 101978
QY 61 GCACCCCTCCATGCTGTGTAACACGATGCTCCCTTACCAAGATTTCTATGAAGA 120
DB 101977 GCACCCCTCCATGCTGTGTAACACGATGCTCCCTTACCAAGATTTCTATGAAGA 101918
QY 121 ACGGGGTTCCTGGAATATTGATGCCCATCATATAGGATTTATCTAAGGGAATCC 180
DB 101917 ATGGGGTTCCTGGAATATTGATGCCCATCATATAGGATTTATCTAAGGGAATCC 101858
QY 181 ACCTTCACTGCCACACCATATGCCCGCAACTGCTATAACTCTGCCACTCTTTGCAATG 240
DB 101857 ACCTTCACTGCCACACCATATGCCCGCAACTGCTATAACTCTGCCACTCTTTGCAATG 101798
QY 241 CATGCAATATCTAATATTGACAGGGGAAATGATTAATCTAGTTGCTCTGGAGGACTT 300
DB 101797 CATGCAATATCTAATATTGACAGGGGAAATGATTAATCTAGTTGCTCTGGAGGACTT 101738
QY 301 GGAGCCACTGCTGTGAGCTTACTTACCCCATACAGTATGCTGATGGGGGTGGAATT 360
DB 101737 GGAGCCACTGCTGTGAGCTTACTTACCCCATACAGTATGCTGATGGGGGTGGAATT 101678

QY 361 CAAGGTCTAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 420
DB 101677 CAAGGTCTAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 101618
QY 421 AGCACTCTAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 479
DB 101617 AGCACTCTAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 101558
QY 480 CCATCTCTGCTGCTGAGGCTATTTAAATACCAACCTCTCTCTGCTCCATAGAGTCTCAGC 539
DB 101557 CCATCTCTGCTGCTGAGGCTATTTAAATACCAACCTCTCTCTGCTCCATAGAGTCTCAGC 101498
QY 540 CCATAACCTCTAATCTGTTGATGCTCTCCCTGCTGCTCTGCTGCTGCTGCTGCTGCT 599
DB 101497 CCATAACCTCTAATCTGTTGATGCTCTCCCTGCTGCTCTGCTGCTGCTGCTGCTGCT 101438
QY 600 CCCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 659
DB 101437 CCCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 101378
QY 660 AGGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 719
DB 101377 AGGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 101318
QY 720 TAGCAATATCTAGACACCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 779
DB 101317 TAGCAATATCTAGACACCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 101258
QY 780 ATAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 839
DB 101257 ATAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 101198
QY 840 GAATGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 899
DB 101197 GAATGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 101139
QY 900 CACTGAAACGAGATTTATACAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 959
DB 101138 CACTGAAACGAGATTTATACAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 101079
QY 960 TCTTCTCTTTGTTTATCAGACGAGAGTGTCTAGGACAGTCTAGGACAGTCTGCTGCTGCTGCTGCT 1019
DB 101078 TCTTCTCTTTGTTTATCAGACGAGAGTGTCTAGGACAGTCTAGGACAGTCTGCTGCTGCTGCT 101019
QY 1020 CACAACCTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1079
DB 101018 CACAACCTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 100959
QY 1080 GGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1139
DB 100958 GGTTCGCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 100899
QY 1140 TCAAAATCGAAGAGCTTTAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1199
DB 100898 TCAAAATCGAAGAGCTTTAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 100839
QY 1200 AGGAGAGAAACGCTG-----TTATTAATGTTAATCA 1229
DB 100838 AGGAGAGAAATGTTGTTTATGTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATCA 100779
QY 1230 ATCCAGAAATGCTCACTGAGAGAGTGTAAAGAAATTCGAGATCGAATACAAATGATAGAGAGA 1289
DB 100778 ATCTGAAATGCTCACTGAGAGAGTGTAAAGAAATTCGAGATCGAATACAAATGATAGAGAGA 100719
QY 1290 GGAGCTTC-AAAAACACCGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1348
DB 100718 GGAGCTTC-AAAAACACCGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 100659
QY 1349 CTTCTTAGGACCTCTAGAGCTCTAATATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1408
DB 100658 CTTCTTAGGATCTCTAGAGCTCTAATATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 100599

QY 1124 TAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTTAGACTTTGCTTAACCGCCAAAAGAGGGG 1183
DB 945 TAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTTAGACTTTGCTTAACCGCGGAAAGCGGG 1004
QY 1184 GAACCTGTTTATTTTATAGGAAGAAACGCTGTTATTTATGTTTAAATCAATCCAGAAATGCTCA 1243
DB 1005 GAACCTTTTATTTTATAGGAAGAAATGCTGTTTATGTTTAAATCAATCCGGAATCATCA 1064
QY 1244 CTGAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACA 1303
DB 1065 CCAGAGAAAGTTAAAGAAATTCAGAGTTCGAATATAACGTAGAGCAAGAGCTGCAAAACA 1124
QY 1304 CCGAAAGCTGGGGCCTCCTCAGGCAATGGAATGCGCTGGGTTCCTCCCTCTTTAGGACCTC 1363
DB 1125 CTGGACCTGGGGCCTCCTCAGCCCAATGGAATGCGCTGGGTTCCTCCCTCTTTAGGACCTC 1184
QY 1364 TAGCAGCTTAATATTTGTTACTCTCTTTGGACCTGTTATCTTTAACCTCTTTGTTAAGT 1423
DB 1185 TAGCAGCTAATATTTGTTACTCTCTTTGGACCTGTTATCTTTAACCTCTTTGTTAAGT 1244
QY 1424 TTCTCTCTCCAGAAATGAAGCTGTAAGCTACAGATGGTCTTACAAATGGAACCCCA 1481
DB 1245 TTGTCTTTCCAGAAATCGAAGCAGTAAACTACAAATCGTTCCTTCAAAATGGAGCCCA 1302

RESULT 16

US-10-632-793-24
; Sequence 24, Application US/106322793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glauclia
; APPLICANT: MALLET, Francois
; APPLICANT: VOISSET, Cecile
; TITLE OF INVENTION: ENDOGENOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
; FILE OF INVENTION: AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
; FILE REFERENCE: 110048
; CURRENT APPLICATION NUMBER: US/10/632.793
; PRIOR FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: US/09/869,927
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: PCT/FR00/00144
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: FR 99/00888
; PRIOR FILING DATE: 1999-01-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 1948
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84)..(84)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (193)..(193)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (241)..(241)
; OTHER INFORMATION: n = a or g or c or t/u
US-10-632-793-24

Query Match 75.4%; Score 1116.4; DB 16; Length 1948;
Best Local Similarity 94.2%; Pred. No. 0;
Matches 1159; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 1 ATGGCCCTCCCTTATCATATCTTTCTTTACTGTTGTTTACCCCTCTTTCACCTCTCTCACT 60
DB 719 ATGGCCCTCCCTTATCATATCTTTCTTTACTGTTGTTTACCCCTCTTTCACCTCTCTCACT 778

QY 61 GCACCCCTCCATGCTGCTGTATACACCAAGTAGTCCCTTACCAGAGTTTCTTATGAAGA 120
DB 779 GCACCCCTCCATGCTGCTGTATACCAAGTAGTCCCTTACCAGAGTTTCTTATGAAGA 838
QY 121 AGCGGCTTCCCTGGAATAATTGATGCCCATCATATAGGAGTTTATCTTAAGGGAACCTCC 180
DB 839 ATGCGAGCTCCCGGAATAATTGATGCCCATCATATAGGAGTTTATCTTAAGGGAACCTCC 898
QY 181 ACCTTCACTGCCACACCCATATGCGCAACTGCTATCACTCTGCGCACTCTTTGCGATG 240
DB 899 ACCTTCACTGCCACACCCATATGCGCAACTGCTATCACTCTGCGCACTCTTTGCGATG 958
QY 241 CATGCAATACTCATTTATTTGGACAGGAAATGATTAATCTCTAGTTGCTCTGAGACTT 300
DB 959 CATGCAATACTCATTTATTTGGACAGGAAATGATTAATCTCTAGTTGCTCTGAGACTT 1018
QY 301 GGAGCCACTGCTGTTGGACTTACTTACCCATACCCAGTATGCTGATGGGGTGGAAAT 360
DB 1019 GGAGTCACTGCTGTTGGACTTACTTACCCAACTGGTATGCTGATGGGGTGGAGTT 1078
QY 361 CAAGGTTCAGGCAAGAGAAAAAACAAGTAAAGAGCAATCTCCCAACTGACCCGGGACAT 420
DB 1079 CAAGATCAGGCAAGAGAAAAAACAAGTAAAGAGCAATCTCCCAACTGACCCGGGACAT 1138
QY 421 AGCACCCCTAGCCCTTACAAAGGACTAGTTCTCTCAAACTACATGAACCCCTCCGTACC 480
DB 1139 GGCACCTCTAGCCCTTACAAAGGACTAGATCTCTCAAACTACATGAACCCCTCCGTACC 1198
QY 481 CATACTCGCTGCTGAGCCCTATTTAATACCACTCCTCACTCGGCTCCTCAGGCTCTCAGCC 540
DB 1199 CATACTCGCTGCTGAGCCCTATTTAATACCACTCCTCACTCGGCTCCTCAGGCTCTCAGCC 1258
QY 541 CAAAACCTCTAACTGTTGGATGTCCTCCCTGCTCCTCCTGCTCCTGAGGCATATGTTTCAATC 600
DB 1259 CAAAACCTCTAACTGTTGGATGTCCTCCCTGCTCCTCCTGCTCCTGAGGCATATGTTTCAATC 1318
QY 601 CCGTCTCTTGAACAATGGAACAACCTTTCAGCAGAGAAATAAACCCTTCCCGTTTATGTA 660
DB 1319 CCGTCTCTTGAACAATGGAACAACCTTTCAGCAGAGAAATAAACCCTTCCCGTTTATGTA 1378
QY 661 GGACCTCTTGTTCCTCAATCTGGAATAAACCCTTCACTCCTCAACCTCCTGTTGTAATAATTT 720
DB 1379 GGACCTCTTGTTCCTCAATCTGGAATAAACCCTTCACTCCTCAACCTCCTGTTGTAATAATTT 1438
QY 721 AGCAATATATAGACAACAACCTGCTCCCAATGCAATGCAATGCAATGCAATGCAATGCAAT 780
DB 1439 AGCAATATATAGACAACAACCTGCTCCCAATGCAATGCAATGCAATGCAATGCAATGCAAT 1498
QY 781 ATAGTCTGCTACCTCCCTCAGGAATATTTTGTCTGTGTGCTACCTCAGCTATCATTTGTTTG 840
DB 1499 ATAGTCTGCTACCTCCCTCAGGAATATTTTGTCTGTGTGCTACCTCAGCTATCATTTGTTTG 1558
QY 841 AATGGCTCTTCAAGATCTATGCTCTTCTCTCAATCTTATGTCCTCCCTATGACATCTAC 900
DB 1559 AATGGCTCTTCAAGATCTATGCTCTTCTCTCAATCTTATGTCCTCCCTATGACATCTAC 1618
QY 901 ACTGAACAAGATTTATCAATCATGCTGCTACCTAAGCCCCCAACAAGAGTAGTACCCATT 960
DB 1619 ACTGAACAAGATTTATACAGTTATGTCATCTAAGCCCCCAACAAGAGTAGTACCCATT 1678
QY 961 CTTCTCTTTGTTATCAGAGCAGGAGTGTAGGAGACTAGGTACTGTCATTTGGCAGTATC 1020
DB 1679 CTTCTCTTTGTTATAGAGCAGGAGTGTAGGAGACTAGGTACTGTCATTTGGCAGTATC 1738
QY 1021 ACAACCTCTACTCAGTTTCTACTACAACTATCTCAAGAAATAAATGTTGTCATGGAACAG 1080
DB 1739 ACAACCTCTACTCAGTTTCTACTACAACTATCTCAAGAAATAAATGTTGTCATGGAACAG 1798
QY 1081 GTACGTGACTCCCTGGTCACTTTCAGAGATCACTTAACCTCCCTAGCAGCAGTAGTCTCT 1140
DB 1799 GTGCGGACTCCCTGGTCACTTTCAGAGATCACTTAACCTCCCTAGCAGCAGTAGTCTCT 1858
QY 1141 CAAAATCGAGAGCTTTAGACTTTGCTAAACCCCAAGAGGGGGAACCTCTTTATTTTATA 1200

Db 1859 CAAATCGAAGAGCTTTAGACTTGTCTAAACCGCTGAAGAGGGGGAACCTGTTTATTTTAA 1918
Qy 1201 GGAGAAGAACGCTGTTATTAATGTTAAATCAA 1230
Db 1919 GCGGAAGAATGCTGTTATTATGTTAAATCAA 1948

RESULT 17

US-09-864-761-21192
; Sequence 21192, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21192
; LENGTH: 792
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002346.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
; OTHER INFORMATION: NT HIT: AF208161.1, EVALUATE 0.00e+00
; OTHER INFORMATION: EST HUMAN HIT: AU138405.1, EVALUATE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P10269, EVALUATE 2.00e-16
US-09-864-761-21192

Query Match 48.4%; Score 716.6; DB 9; Length 792;
Best Local Similarity 94.7%; Pred. No. 1e-209;
Matches 753; Conservative 0; Mismatches 39; Indels 3; Gaps 1;

Qy 375 AGAAAACAAGTAAAGGAAAGCAATCTCCCAACTGACCGGGGACATAGCAGCCCTAGCCCC 434
Db 1 AGAAAACAACATTAAGGAAGTAAATCTCCCAACTGACCTGGGTACATAGCAGCCCTAGCCCC 60
Qy 435 CTACAAAGGACTAGTGTCTCTCAAAACTACATGAAGCCCTCCGCTACCCATACCTGCGCTGGT 494
Db 61 CTACAAAGGACTAGTGTCTCTCAAAACTACATGAAGCCCTCCATACCCATACCTGCGCTGGT 120
Qy 495 GAGCCTATTATTAATACCAACCTCTACCTGGCTTCATAGAGGTCTCAGCCCAAAACCCCTACTAA 554
Db 121 AAGCCTATTATTAATACCAACCTCTACCTGGCTTCATAGAGGTCTCAGCCCAAAACCCCTACTAA 180
Qy 555 CTGTTGGATGTGCTCCCTCCCTGACCTTCAGGCCATACATTTCAATCCCTGTTCTTGAACA 614
Db 181 CTGTTGGATGTGCTCCCTCCCTGACCTTTAGGCCATACATTTCAATCCCTGTTCTTGAACA 240
Qy 615 ATGGAACAACCTTCAGCACAGAAATAAACACCACTTCGCTTTTAGTAGGACCTCTTGTTTC 674
Db 241 ATGGAACAACCTTCAGCACAGAAATAAACACCACTTCGCTTTTAGTAGGACCTCTTGTTTC 297
Qy 675 CAATCTGGAATAAACCCATACCTCAAACTCCTCAAACTCCTGTAATAATTTAGCAATACTATAGA 734
Db 298 CAATCTGGAATAAACCCATACCTCAAACTCCTCAAACTCCTGTAATAATTTAGCAATACTATAGA 357
Qy 735 CACAACGAGTCCCAATGATCAGGTGGGTGAACACCTCCACACGCAATAGTCTGCCTACC 794
Db 358 CACAGCAACCTCCCAATGATCAGGTGGGTGAACCTCCACACGCAATAGTCTGCCTACC 417
Qy 795 CTCAGGAATATTTTGTCTGTGGTACCTCAGGCTATCATTTGTTGAATGGCTCTTCAGA 854
Db 418 CTCAGGAATATTTTGTCTGTGGTACCTCAGGCTATCATTTGTTGAATGGCTCTTCAGA 477
Qy 855 ATCTATGTCTCTCTCTCAATTTAGTGCCCTCTATGACCATCTACACTGAAACAAGATT 914
Db 478 ATCTGTGTGCTCTCTCTCAATTTAGTGCCCTCTATGACCATCTACACTGAAACAAGATT 537
Qy 915 ATCAATCATGTGCTACCTTAAGCCCCCAACAAGAGTAGTACCCATTTCTCTTTTGTAT 974
Db 538 ATCAATCATGTGCTACCTTAAGCCCCCAACAAGAGTAGTACCCATTTCTCTTTTGTAT 597
Qy 975 CAGAGCAGGAGTGCTAGGAGACTAGGTACTGGCATTGGCAGTATCACAACTCTACTCA 1034
Db 598 TGGAGCAGGAGTGCTAGGAGAGTAGTACTGGCATTGGCAGTATCACAACTCTACTCA 657
Qy 1035 GTTCTACTACAACTATCTCAAGAAATAAATGGTGTGATGGAAACAGGTCTACTGACCTCCCT 1094
Db 658 GTTCTACTACAACTATCTCAAGAAATAAATGGTGTGATGGTGGTGGTGGTGGTGGTGGTGGT 717
Qy 1095 GGTACCTTTGCAAGATCAACTTAACCTCCCTAGCAGCAGTAGTCTCTTCAAAAATCGAAGAGC 1154
Db 718 GGTACCTTTGCAAGATCAACTTAACCTCCCTAGCAGCAGTAGTCTCTTCAAAAATCGAAGAGC 777
Qy 1155 TTTAGACTTGTAAAC 1169
Db 778 TTTAGACTTGTAAAC 792

RESULT 18

US-10-363-616-228
; Sequence 228, Application US/10363616
; Publication No. US20040044181A1

GENERAL INFORMATION:
; APPLICANT: Hiseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-113 (793)
; CURRENT APPLICATION NUMBER: US/10/363,616
; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: 09/654,935
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 490
; SEQ ID NO 228
; LENGTH: 1684
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (89)..(1684)
US-10-363-616-228

Query Match 40.0%; Score 591.8; DB 16; Length 1684;
Best Local Similarity 98.8%; Pred. No. 4.8e-171;
Matches 596; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 879 AGTGGCCCTATGACCATCTACACTGACGACGAGATTATACAATCATGTCGTACTAAGCC 938
DB 721 AGTGGCCCTATGACCATCTACACTGACGACGAGATTATACAATCATGTCGTACTAAGCC 780
QY 939 CCACAAAGAGAGTACCCATTCTTCCTTTTGTATCAGAGCAGGAGTGTAGGCAGACT 998
DB 781 CCACAAAGAGAGTACCCATTCTTCCTTTTGTATCAGAGCAGGAGTGTAGGCAGACT 840
QY 999 AGTACTGGCATTGGCAGTATCAACCTCTACTCAGTCTCTAACAATCTCAAGA 1058
DB 841 AGTACTGGCATTGGCAGTATCAACCTCTACTCAGTCTCTAACAATCTCAAGA 900
QY 1059 AATAATGGTGACATGGACAGCTCACTGACTCCCTGGTCACTGCTCAAGATCAACTAA 1118
DB 901 AATAATGGTGACATGGACAGCTCACTGACTCCCTGGTCACTGCTCAAGATCAACTAA 960
QY 1119 CTCCTAGCAGCAGTAGTCTCTCAAAATCGAAGAGCTTTAGACTTGTAAACCCCAAAAG 1178
DB 961 CTCCTAGCAGCAGTAGTCTCTCAAAATCGAAGAGCTTTAGACTTGTAAACCCCAAAAG 1020
QY 1179 AGGGGAACCTGTTATTTTATAGAGAGAAACGCTGTTATATGTTAAATCAATCCAGAT 1238
DB 1021 AGGGGAACCTGTTATTTTATAGAGAGAAATGCTGTTATATGTTAAATCAATCCAGAT 1080
QY 1239 TGTCACTGAGAAATTAAGAAATTCGAGATCGAATACATATAGCAGGAGGCTTCA 1298
DB 1081 TGTCACTGAGAAATTAAGAAATTCGAGATCGAATACATATAGCAGGAGGCTTCA 1140
QY 1299 AAACACCGAACGCTGGGGCCCTCTCAGCCAAATGGATGCCCTGGGTCTCCCTCTTAGG 1358
DB 1141 AAACACCGAACGCTGGGGCCCTCTCAGCCAAATGGATGCCCTGGGTCTCCCTCTTAGG 1200
QY 1359 ACCTTAGCAGCTCTAATATGTTACTCTCTTTTGGACCCGCTGATCTTTAACTCTCTGT 1418
DB 1201 ACCTTAGCAGCTCTAATATGTTACTCTCTTTTGGACCCGCTGATCTTTAACTCTCTGT 1260
QY 1419 TAAGTTGCTCTTCCAGAAATTCAGCTGTAAAGCTACAGATGGTCTTACAAATGGACC 1478
DB 1261 TAAGTTGCTCTTCCAGAAATTCAGCTGTAAAGCTACAGATGGTCTTACAAATGGACC 1320
QY 1479 CCA 1481
DB 1321 CCA 1323

RESULT 19

US-10-632-793-19
; Sequence 19, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia

RESULT 20

US-10-632-793-25
; Sequence 25, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: MALLET, Francois

APPLICANT: MALLET, Francois
APPLICANT: VOISSET, Cecile
TITLE OF INVENTION: ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
TITLE OF INVENTION: AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
FILE REFERENCE: 110048
CURRENT APPLICATION NUMBER: US/10/632,793
CURRENT FILING DATE: 2003-08-04
PRIOR APPLICATION NUMBER: US/09/869,927
PRIOR FILING DATE: 2001-10-22
PRIOR APPLICATION NUMBER: PCT/FR00/00144
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: FR 99/00888
PRIOR FILING DATE: 1999-01-21
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.1
SEQ ID NO 19
LENGTH: 591
TYPE: DNA
ORGANISM: Homo sapiens
US-10-632-793-19

Query Match 39.4%; Score 582.8; DB 16; Length 591;
Best Local Similarity 99.7%; Pred. No. 1.6e-168;
Matches 584; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 887 CTATGACCCTACACTGAAACAAGATTTATACAATCATGTCGTACTAAGCCCAACA 946
DB 1 CCATGCCCATCTACACTGAAACAAGATTTATACAATCATGTCGTACTAAGCCCAACA 60
QY 947 AAAGAGTACCCATTCTTCCTTTTGTATCAGAGCAGGAGTGTAGCAGACTAGGTACTG 1006
DB 61 AAAGAGTACCCATTCTTCCTTTTGTATCAGAGCAGGAGTGTAGCAGACTAGGTACTG 120
QY 1007 GCATTGGCAGTATCAAACTCTACTCAGTCTTACTACAAAATCTCTCAAGAAATAAATG 1066
DB 121 GCATTGGCAGTATCAAACTCTACTCAGTCTTACTACAAAATCTCTCAAGAAATAAATG 180
QY 1067 GTGACATGGAACAGGTCCTGACTCCCTGGTCACTGCAAGATCAACTTAATCCCTAG 1126
DB 181 GTGACATGGAACAGGTCCTGACTCCCTGGTCACTGCAAGATCAACTTAATCCCTAG 240
QY 1127 CAGCAGTACTCTTCAAAATCGAAGCTTTAGACTTGTAAACCCCAAAAGAGGGGAA 1186
DB 241 CAGCAGTACTCTTCAAAATCGAAGCTTTAGACTTGTAAACCCCAAAAGAGGGGAA 300
QY 1187 CTTGTTTATTTAGAGAGAAACGCTGTTATGTTATGTTAATCAATCCAGAAATGTCAGTG 1246
DB 301 CTTGTTTATTTAGAGAGAAACGCTGTTATGTTAATCAATCCAGAAATGTCAGTG 360
QY 1247 AGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGACAGAGGAGCTTCAAAACACCG 1306
DB 361 AGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGACAGAGGAGCTTCAAAACACCG 420
QY 1307 AACGCTGGGGCCCTCTCAGCCAAATGGATGCCCTGGGTCTCCCTCTTTAGGACCTCTAG 1366
DB 421 AACGCTGGGGCCCTCTCAGCCAAATGGATGCCCTGGGTCTCCCTCTTTAGGACCTCTAG 480
QY 1367 CAGCTCTAATATGTTACTCTCTTTGGACCCCTGTATCTTTAACTCTCTTTGTTAAGTTTG 1426
DB 481 CAGCTCTAATATGTTACTCTCTTTGGACCCCTGTATCTTTAACTCTCTTTGTTAAGTTTG 540
QY 1427 TCTCTTCCAGAAATTAAGCTGTAAAGCTACAGATGGTCTTACAAAT 1472
DB 541 TCTCTTCCAGAAATTAAGCTGTAAAGCTACAGATGGTCTTACAAAT 586


```

/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ POLYMORPHISMS IN THE HUMAN GENOME
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ CURRENT FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 322491
/ LENGTH: 570
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(570)
/ OTHER INFORMATION: n = A,T,C or G
US-10-027-632-322491

```

Query Match	25.58;	Score 377.8;	DB 13;	Length 570;
Best Local Similarity	90.7%;	Pred. No. 2.6e-105;		
Matches 411;	Conservative 2;	Mismatches 39;	Indels 1;	Gaps 1;
QY	1	ATGCGCCCTCCCTTATCATACATCTTTTCTCTTATCTGTGTTCTTATACCCCTTTGGCTCTCACT	60	
Db	463	ATAGCCCTCCCTTATCATATTTTCTCTTTACCATTTCTTTACCTGTTTTCACCTCACT	404	
QY	61	GCACCCCTTCATGCTGCTGTACAAACAGTAGCTCCCTTACCAAGAGTTTCTATGAAGA	120	
Db	403	TCACCTCACTCCATGCTGTYGTATGACAGTAGTCCCTTACCAAGAGTTTCT-TGGAGA	345	
QY	121	ACGGGGTTTCTTGGAAATATTGATGCCCCCATCATATAGGAGTTTATCTAAGGGAAATCC	180	
Db	344	ATGCAGTTTCTTGGAAATATTGATTTCCCATCATATAGGAGTTTATCTAAGGGAAATCAC	285	
QY	181	ACTTTCACTGCCACACCCATATGCCCGGCAATGCTATTAACCTGCGCACTTTTGCATG	240	
Db	284	ACTTTCACTGCCACACCCATATGCCCTACAACCTGCTATAACGTRCCACTTTTTCATG	225	
QY	241	CATGCAAAATACTCATTTATGACAGGGAAATGATTAATCCTAGTTTCTCTGGAGGACTT	300	
Db	224	CATGCAAAATACTAATTTATTTGTTTCATGGAAATGATTAATCCTTGTGTCTGGAGGACTT	165	
QY	301	GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCATATGTTCTGATGGGGTGGAAATT	360	
Db	164	GGAGCCACTGTCTGTTGGACTTACTTCACCCATACCATATGTTCTGATGGGGTGGAGTT	105	
QY	361	CAAGGTCAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCCGGGACAT	420	
Db	104	CAAGATCAGGCAAGAGAAAAACAAGTAAAGGAAGTAATCTCCCAACTGACCCCTGGTACAT	45	
QY	421	AGCACCCCTTAGCCCTTACAAAGGACTAGTTCTC	453	
Db	44	AGCACCCCTTAGACCCCTACAAAGGACTAATCCCC	12	

RESULT 23
US-10-027-632-322492/c
; Sequence 322492, Application US/10027632
: Publication No. US20020198371A1

```

; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 322492
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(570)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-322492

```

Query Match	25.58;	Score 377.8;	DB 13;	Length 570;
Best Local Similarity	90.74;	Pred. No. 2.6e-105;		
Matches 411;	Conservative 2;	Mismatches 39;	Indels 1;	Gaps 17
Qy	1	ATGGCCCTCCCTTATCATACTTTTCTCTTTACTGTGTTCTTTACCCCTTTTCGCTCTCACT	60	
Db	463	ATAGCCCTCCCTTATCATATTTTCTCTTTACCAATCTCTTACCTGCTTTCACTCTCACT	404	
Qy	61	GCACCCCTTCATGCTGCTGTACACACAGTAGCTCCCTTACCAAGAGTTCTCTATGAGA	120	
Db	403	TCACCTCACTCCATGCTGTGTATGACACAGTAGCTCCCTTACCAAGAGTTTCT-TGGAGA	345	
Qy	121	ACGGCGTTCTCTGGAAATATTGTATGCCCCCATCATATAGGAGTTTATCTAAGGGAAATCC	180	
Db	344	ATGCAGTTCTCTGGAAATATTGTATTCCTCCCATCATATAGGAGTTTATCTAAGGGAAATCAC	285	
Qy	181	ACCTTCACTGCCACACCAATATGCCCGCACTGTCTATAACTCTGCCACTCTTTTGCATG	240	
Db	284	ACCTTCACTGCCACACCAATATGCCCTCACACTGCTATAACGTCRCACCTTTTTCATG	225	
Qy	241	CATGCAAACTACTCATATTGCGACAGGGAAAATGATTAACTCTAGTTGTCTGGAGGACTT	300	
Db	224	CATGCAAACTACTAATTATTGTTTCATGGAATGATTAACTCTTGTCTGGAGGACTT	165	
Qy	301	GGAGCCACTGTCTGTGTTGGACTTACTTCACCCATACCAAGTATGTCGTGATGGGGGTGGAATT	360	
Db	164	GGAGCCACTGTCTGTGTTGGACTTACTTCACCCCTACCAAGTATGTCGTGATGGGGGTGGAGTT	105	
Qy	361	CAAGETCAGCGCAGAGAAAAACAAGTAAAGAAGCAATCTCCCAACTGACCCCGGGACAT	420	
Db	104	CAAGATCAGCGCAGAGAAAAACAAGTAAAGAAGTAACTCCCAACTGACCCCTGGTACAT	45	
Qy	421	AGCACCCCTTAGCCCTTACAAAGGACTAGTTCTC	453	
Db	44	AGCACCCCTTAGACCTCAAGAGGACTAATCCCC	12	

RESULT 24
US-10-027-632-322491/c
: sequence 322491, Application US/10027632

```
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 322491
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc. feature
; LOCATION: (1)...(570)
; OTHER INFORMATION: n = A,T,C or G
; US-10-027-632-322491

Query Match      25.5%; Score 377.8; DB 15; Length 570;
Best Local Similarity 90.7%; Pred. No. 2.6e-105;
Matches 411; Conservative 2; Mismatches 39; Indels 1; Gaps 1;

Qy 1 ATGCCCCCTCCCTTATCATACATCTTTCTCTTTACTGTTCTCTTACCCCTTCCTTACCCCTTCGCTCTCACT 60
Db 463 ATAGCCCTCCCTTATCATATTTTCTCTTTTACCATTTCTTTTACCATTTCTTACCTGCTTTCACTCTCACT 404

Qy 61 GCACCCCTCTCATGCTGCTGTACAAACAGTAGTCTCCCTTACCAAGAGTTTCTATGAAGA 120
Db 403 TCACCTCACTCCATGCTGTGTATGACCAAGTAGTCTCCCTTACCAAGAGTTTCT-TGGAGA 345

Qy 121 ACGCGGCTTCTGGAATATTGATGCCCATCATATAGGAGTTTATCTAAGGGAACTCC 180
Db 344 ATGCAGCTTCTGGAATATTGATTTCCCATCATATAGGAGTTTATCTAAGGGAAATCAC 285

Qy 181 ACCTTCACTGCCACACCCCATATGCCCCCGCAACTGCTATTAACCTGCGCACTCTTTGCAATG 240
Db 284 ACCTTCACTGCCACACCCCATATGCCCCCATATGCGCTACAACTGCTATTAACGCTTTCATG 225

Qy 241 CATGCAAAATCTCATTTATGACAGGGAATGATTAATCTCTAGTTGCTCTGGAGGACTT 300
Db 224 CATGCAAAATCTAATTTATTTGTCATGGAAAATGATTAATCTCTGCTCTGGAGGACTT 165

Qy 301 GGAGCCACTGCTCTGGAGCTTACTTCAACCATACAGTAGTATGCTGATGGGGTGGAAAT 360
Db 164 GGAGCCACTGCTCTGGAGCTTACTTCAACCCCTACCAAGTAGTATGCTGATGGGGTGGAGTT 105

Qy 361 CAAGGTGAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGACAT 420
Db 104 CAAGATCAGGCAAGAGAAAAACAAGTAAAGGAAGTATCTCCCAACTGACCCCTGTATCAT 45

Qy 421 AGCACCCCTAGCCCTTACAAAGGACTAGTTTCTC 453
Db 44 AGCACCCCTAGACCTTACAAAGGACTAATCCCC 12

RESULT 25
US-10-027-632-322492/c
```

```
; Sequence 322492, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 322492
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc. feature
; LOCATION: (1)...(570)
; OTHER INFORMATION: n = A,T,C or G
; US-10-027-632-322492

Query Match      25.5%; Score 377.8; DB 15; Length 570;
Best Local Similarity 90.7%; Pred. No. 2.6e-105;
Matches 411; Conservative 2; Mismatches 39; Indels 1; Gaps 1;

Qy 1 ATGCCCCCTCCCTTATCATACATCTTTCTCTTTACTGTTCTCTTACCCCTTCCTTACCCCTTCGCTCTCACT 60
Db 463 ATAGCCCTCCCTTATCATATTTTCTCTTTTACCATTTCTTTTACCATTTCTTACCTGCTTTCACTCTCACT 404

Qy 61 GCACCCCTCTCATGCTGCTGTACAAACAGTAGTCTCCCTTACCAAGAGTTTCTATGAAGA 120
Db 403 TCACCTCACTCCATGCTGTGTATGACCAAGTAGTCTCCCTTACCAAGAGTTTCT-TGGAGA 345

Qy 121 ACGCGGCTTCTGGAATATTGATGCCCATCATATAGGAGTTTATCTAAGGGAACTCC 180
Db 344 ATGCAGCTTCTGGAATATTGATTTCCCATCATATAGGAGTTTATCTAAGGGAAATCAC 285

Qy 181 ACCTTCACTGCCACACCCCATATGCCCCCGCAACTGCTATTAACCTGCGCACTCTTTGCAATG 240
Db 284 ACCTTCACTGCCACACCCCATATGCCCCCATATGCGCTACAACTGCTATTAACGCTTTCATG 225

Qy 241 CATGCAAAATCTCATTTATGACAGGGAATGATTAATCTCTAGTTGCTCTGGAGGACTT 300
Db 224 CATGCAAAATCTAATTTATTTGTCATGGAAAATGATTAATCTCTGCTCTGGAGGACTT 165

Qy 301 GGAGCCACTGCTCTGGAGCTTACTTCAACCATACAGTAGTATGCTGATGGGGTGGAAAT 360
Db 164 GGAGCCACTGCTCTGGAGCTTACTTCAACCCCTACCAAGTAGTATGCTGATGGGGTGGAGTT 105

Qy 361 CAAGGTGAGGCAAGAGAAAAACAAGTAAAGGAAGCAATCTCCCAACTGACCCGGGACAT 420
Db 104 CAAGATCAGGCAAGAGAAAAACAAGTAAAGGAAGTATCTCCCAACTGACCCCTGTATCAT 45

Qy 421 AGCACCCCTAGCCCTTACAAAGGACTAGTTTCTC 453
Db 44 AGCACCCCTAGACCTTACAAAGGACTAATCCCC 12

RESULT 26
```

```

RESULT 28
US-10-027-632-322575
; Sequence 322575, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006

```

; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 322575
; LENGTH: 551
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(551)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-322575

Query Match 23.8%; Score 353; DB 13; Length 551;
Best Local Similarity 92.0%; Pred. No. 1.1e-97;
Matches 412; Conservative 2; Mismatches 30; Indels 4; Gaps 4;
QY 1 ATGGCCCTCCCTTATCATATCTTTCTTTTACGTGTTCTTACCCCTTTGGCTCTCACT 60
Db 102 ATGGCCCTCCCTTATCATATCTTTCTTTTACGTGTTCTTACCCCTTTGGCTCTCACT 161
QY 61 GCACCCCTCCCTGCTGTACCAACAGTAGTCCCTTACCAAGAGTTTCTATGAAGA 120
Db 162 GCACCCCTCCCTGCTGTACCAACAGTAGTCCCTTACCAAGAGTTTCTATGAGA 221
QY 121 ACGGCTCTCTGGAAATATTGATGCCCATCATATAGAGTTTATTAAGGAACTCC 180
Db 222 ATGGCTCTCTGGAAATATTGATGCCCATCATATAGAGTTTATTAAGGAACTCC 280
QY 181 ACCTTCACTGCCACACCCATATGCCCGCACTGCTATAACTCTGCCACTCTTTCGATG 240
Db 281 ACCTTCACTGCCACACCCATATGCCCGCACTGCTATAACTCTGCCACTCTTTCGATG 340
QY 241 CATGCAATACTCATTTATTGGACAGGAAATGATTAACTCTAGTTGTCTCGAGGACTT 300
Db 341 CATGCAATACTCATTTATTGGACAGGAAATGATTAACTCTAGTTGTCTCGAGGACTT 400
QY 301 GGAGCCACTGTCTGTGGACTT-ACCTTCACTACCACTACAGTATGCTGTGATGGGGTGAAT 359
Db 401 GGAGCCACTGTCTGTGGACTTAACTTCACTATACCGGTATGCTGTGATGGGGTGAAT 460
QY 360 TCAAGTTCAGGCAAGGAAACAACTAAGGAAGCAATCTCCCACTGACCCGGGACA 419
Db 461 TC-AGATGAGGCAAGAG-AAAAACGTANAGGAAGTAAATCTCCANCTGACNGTGTGACA 518
QY 420 TAGCACCCCTAGCCCTTACAAAGGACTA 447
Db 519 TAGCACCCCTTACCCCTTACAAAGGACTA 546

RESULT 29
US-10-027-632-322574
; Sequence 322574, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108927.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 322574
; LENGTH: 551
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(551)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-322574

Query Match 23.8%; Score 353; DB 15; Length 551;
Best Local Similarity 92.0%; Pred. No. 1.1e-97;
Matches 412; Conservative 2; Mismatches 30; Indels 4; Gaps 4;
QY 1 ATGGCCCTCCCTTATCATATCTTTCTTTTACGTGTTCTTACCCCTTTGGCTCTCACT 60
Db 102 ATGGCCCTCCCTTATCATATCTTTCTTTTACGTGTTCTTACCCCTTTGGCTCTCACT 161
QY 61 GCACCCCTCCCTGCTGTACCAACAGTAGTCCCTTACCAAGAGTTTCTATGAAGA 120
Db 162 GCACCCCTCCCTGCTGTACCAACAGTAGTCCCTTACCAAGAGTTTCTATGAGA 221
QY 121 ACGGCTCTCTGGAAATATTGATGCCCATCATATAGAGTTTATTAAGGAACTCC 180
Db 222 ATGGCTCTCTGGAAATATTGATGCCCATCATATAGAGTTTATTAAGGAACTCC 280
QY 181 ACCTTCACTGCCACACCCATATGCCCGCACTGCTATAACTCTGCCACTCTTTCGATG 240
Db 281 ACCTTCACTGCCACACCCATATGCCCGCACTGCTATAACTCTGCCACTCTTTCGATG 340
QY 241 CATGCAATACTCATTTATTGGACAGGAAATGATTAACTCTAGTTGTCTCGAGGACTT 300
Db 341 CATGCAATACTCATTTATTGGACAGGAAATGATTAACTCTAGTTGTCTCGAGGACTT 400
QY 301 GGAGCCACTGTCTGTGGACTT-ACCTTCACTACCACTACAGTATGCTGTGATGGGGTGAAT 359
Db 401 GGAGCCACTGTCTGTGGACTTAACTTCACTATACCGGTATGCTGTGATGGGGTGAAT 460
QY 360 TCAAGTTCAGGCAAGGAAACAACTAAGGAAGCAATCTCCCACTGACCCGGGACA 419
Db 461 TC-AGATGAGGCAAGAG-AAAAACGTANAGGAAGTAAATCTCCANCTGACNGTGTGACA 518
QY 420 TAGCACCCCTAGCCCTTACAAAGGACTA 447
Db 519 TAGCACCCCTTACCCCTTACAAAGGACTA 546

RESULT 30
US-10-027-632-322575
; Sequence 322575, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108927.129
; CURRENT APPLICATION NUMBER: US/10/027,632

```
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 322575
; LENGTH: 551
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(551)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-322575

Query Match      23.8%; Score 353; DB 15; Length 551;
Best Local Similarity 92.0%; Pred. No. 1.1e-97;
Matches 412; Conservative 2; Mismatches 30; Indels 4; Gaps 4;

Qy      1  ATGCGCCCTCCCTTATCATATCTTTCTTTTACTGTTCTTTACCCCTTTTCGCTCTCACT 60
Db      102 ATGCGCCCTCCCTTATCATATTTTCTTTTACGTTCTTTACCCCTTTTCACTCTCACT 161

Qy      61  GCACCCCTCCCTGCTGTCTACAAACAGTAGTCCCTTACCAAGATTTCTATGAAGA 120
Db      162 GCACCCCTCCCTGCTGTCTACAAACAGTAGTCCCTTACCAAGATTTCTATGRAGA 221

Qy      121 ACGCGGCTTCTGGAATATTTGATGCCCATCATATAGGAGTTTATCTAAGGAACTCC 180
Db      222 ATGCGGCTTCTGGAATATTTGATGCCCATCATATAGGAGTTTATCTAAGGAACTCC 280

Qy      181 ACCTTCACTGCCACACCCCATATGCCCCGCAACTGCTATTAACCTTGCCACTCTTTGCATG 240
Db      281 ACCTTCACTGCCACACCCCATATGCCCCCAAAATGCTGTAACCTTGCCACTCTTTGCATG 340

Qy      241 CATGCAAAATCTCATTTATTTGGACAGGGAATAATGATTAATCCTAGTTGTCTGGAGACTT 300
Db      341 CATGCAAAATCTCATTTATTTGGACAGGGAATAATGATTAATCCTAGTTGTCTGGAGACTT 400

Qy      301 GGAGCCACTGCTGTTGGACTT-ACCTCACCATACCAGTATGCTGATGGGGTGGAA 359
Db      401 GGAGCCACTGCTGTTGGACTTAACTTCACTATACCGGTATGCTGATGGGGTGGAGT 460

Qy      360 TCAAGGTTCAGGCAAGAGAAAACAAAGTAAAGAGCAATCTCCCAACTGACCCGGGACA 419
Db      461 TC-AGATGAGCAAGAG-AAAACAGTANAGGAAGTAAATCTCCANCTGACNGTGGTACA 518

Qy      420 TAGCACCCCTAGCCCTTACAAAGACTA 447
Db      519 TAGCACCCCTAACCCCTACAAAGACTA 546
```

Search completed: January 22, 2005, 20:08:34
Job time : 812.34 secs

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	1333.8	90.1	2500	3	CR622175	CR622175	full-length
2	1333.8	90.1	2716	3	CR613169	CR613169	full-length
3	1333.8	90.1	2748	3	CR605851	CR605851	full-length
4	1333.8	90.1	2748	3	CR625046	CR625046	full-length
5	1333.8	90.1	2749	3	CR617248	CR617248	full-length
C 6	839	56.7	935	5	BX391741	BX391741	full-length
7	813.6	54.9	955	5	BX347952	BX347952	full-length
8	752.2	50.8	931	5	BX389556	BX389556	full-length
9	747.2	50.5	872	5	BX326647	BX326647	full-length
10	741	50.0	903	5	BX408752	BX408752	full-length
C 11	737.2	49.8	883	5	BX408733	BX408733	full-length
12	728.4	49.2	877	5	BX347619	BX347619	full-length
C 13	714	48.2	936	5	BX388766	BX388766	full-length
14	709.8	47.9	921	5	BX408734	BX408734	full-length
C 15	705.8	47.7	844	5	BX347111	BX347111	full-length
16	703.2	47.5	828	5	BX367907	BX367907	full-length
C 17	698.6	47.2	846	5	BX368078	BX368078	full-length
18	696.2	47.0	995	5	BX389657	BX389657	full-length
19	695	46.9	924	5	BX347314	BX347314	full-length
20	647	43.7	723	1	AUI38405	AUI38405	full-length
21	629.4	42.5	631	8	AQ261133	AQ261133	full-length
22	628.6	42.4	702	1	AUI38097	AUI38097	full-length
23	623.8	42.1	924	5	BX409328	BX409328	full-length
24	609.8	41.2	114	2	BE734284	BE734284	full-length

QY	61	GCACCCCTCCATGCTGCTGTACAAACAGTAGTCCCTTACCAAGAGTTTCTATGAAGA	120
Db	590	GCACCCCTCCATGCGCGTGTATGACCAAGTAGTCCCTTACCAAGAGTTTCTATGAGA	649
QY	121	ACGGGCTTCTGGAAATATTGATGCCCCATCATATATAGGAGTTTATCTAAGGAAACTCC	180
Db	650	ATGACGCTCCGGAAATATTGATGCCCATATGATAGGAGTCTTTCTAAGGGAAACCCC	709
QY	181	ACCTTCACTGCCACACCCATATGCCCGCAACTATTAACCTGCGACCTCTTTGCGATG	240
Db	710	ACCTTCACTGCCACACCCATATGCCCGCAACTGCTATCCTGCGACCTCTTTGCGATG	769
QY	241	CATGCAAAATACTCATTTATGGACAGGGAATAATGATTAATCCTAGTTGTCCTGGAGGACTT	300
Db	770	CATGCAAAATACTCATTTATGGACAGGGAATAATGATTAATCCTAGTTGTCCTGGAGGACTT	829
QY	301	GGAGCCACTGCTGTGTGACTTACTTCAACCATACAGTAGTCTGATGGGGTGGAAATT	360
Db	830	GGAGTCACTGCTGTGTGACTTACTTCAACCAACTGCTATGCTGATGGGGTGGAGTT	889
QY	361	CAAGTTCAGGCAAGAGAAACAAGTAAGGAAGCAATCTCCCACTGACCCGGGACAT	420
Db	890	CAAGATCAGGCAAGAGAAACAAGTAAAGAGTAATCTCCCACTACCCGGGTACAT	949
QY	421	AGCACCCCTAGCCCTACAAAGGACTAGTTCTCTCAAACTACATGAAACCTCCCGTACC	480
Db	950	GGCACTCTAGCCCTACAAAGGACTAGATCTCTCAAACTACATGAAACCTCCCGTACC	1009
QY	481	CATATCGCTGGTGAGCTATTATTAACACCATCTACTCGGCTCAATGAGGTCTAGCC	540
Db	1010	CATATCGCTGGTAAAGCTATTTAATACCAACCTCTACTGGGCTCCATGAGGTCTCGGC	1069
QY	541	CAAAACCTTACTACTGTTGGATGTGCTCCCTGCACTTCAAGGCATACATTTCATC	600
Db	1070	CAAAACCTTACTACTGTTGGATGTGCTCCCTGCACTTCAAGGCATATGTTTCATC	1129
QY	601	CCTGTTCTCTGAAACATGGAACAACTTCAGCACAGAAATAAACACCTTCGTTTGTAGTA	660
Db	1130	CCTGTACTGTAACATGGAACAACTTCAGCACAGAAATAAACACCTTCGTTTGTAGTA	1189
QY	661	GGACCTCTGTTTCCAAATCTGGAATAACCCATACCTCAAACTCAGCTGTGTAATTT	720
Db	1190	GGACCTCTGTTTCCAAATCTGGAATAACCCATACCTCAAACTCAGCTGTGTAATTT	1249
QY	721	AGCAATCTATAGACAAACAGCTCCCAATGATCAGGTGGGTAAACCTCCACACGA	780
Db	1250	AGCAATCTATAGACAAACAGCTCCCAATGATCAGGTGGGTAACTCTCTCCACACAA	1309
QY	781	ATAGTCTGCCTACCTCAGGAATATTTTGTCTGTGTGTACTCAGCCTATCATTTGTTG	840
Db	1310	ATAGTCTGCCTACCTCAGGAATATTTTGTCTGTGTGTACTCAGCCTATCATTTGTTG	1369
QY	841	AATGGCTCTTCAGAACTATGTTGCTTCTCATTTAGTGCCCTCATGACCATCTAC	900
Db	1370	AATGGCTCTTCAGAACTATGTTGCTTCTCATTTAGTGCCCTCATGACCATCTAC	1429
QY	901	ACTGAACAGATTTATACAATCATGCTGCTACTAAGCCCAACAAAGAGTACCCATT	960
Db	1430	ACTGAACAGATTTATACGTTATGTCATATCTAAGCCCAACAAAGAGTACCCATT	1489
QY	961	CTTCTCTTTGTTATCAGAGCAGGAGTGTAGGCACTAGTACTGTCATTGGCAGTATC	1020
Db	1490	CTTCTCTTTGTTATAGGAGCAGGAGTGTAGGCACTAGTACTGTCATTGGCAGTATC	1549
QY	1021	ACAACCTCTACTAGTTCTACTACAATATCTCAAGAAATAAATGATGATGGAACAG	1080
Db	1550	ACAACCTCTACTAGTTCTACTACAATATCTCAAGAAATAAATGATGGAACAG	1609
QY	1081	GTCACTGACTCCCTGGTCACTTGAAGATCAACTTAACCTCCCTAGCAGCAGTAGTCTT	1140
Db	1610	GTGCGGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1669

QY	1141	CAAAATCGAAGAGCTTTAGACTTGTCTAACCAGCAAAAGAGGGGAACCTGTTTATTTTA	1200			
Db	1670	CAAAATCGAAGAGCTTTAGACTTGTCTAACCAGCAAAAGAGGGGAACCTGTTTATTTTA	1729			
QY	1201	GGAGAGAACCGCTGTTTATTTATCTTAATCAATCCAGATTTGTCTACAGAGAAAGTAAAGAA	1260			
Db	1730	GGGGAAGAATGCTGTTTATTTATCTTAATCAATCCGGAATGCTCACTGAGAAGTTAAAGAA	1789			
QY	1261	ATTTCAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAAACACCGAGCGCTGGGCGCTC	1320			
Db	1790	ATTTCAGATCGAATACAACGTAGAGCAGAGGAGCTTCAAAACACCGAGCGCTGGGCGCTC	1849			
QY	1321	CTCAGCAATGAGTCCCTGGGTTCTCCCTCTTTAGGACCTCTAGCAGCTCTAATATTG	1380			
Db	1850	CTCAGCAATGAGTCCCTGGGTTCTCCCTCTTTAGGACCTCTAGCAGCTCTAATATTG	1909			
QY	1381	TTACTCTCTTTGGACCTGATCTTTAACTCTCTTTAAAGTTTGTCTTCCAGAAAT	1440			
Db	1910	CTACTCTCTTTGGACCTGATCTTTAACTCTCTTTAACTCTTTGTTTAACTTGTCTTCCAGATC	1969			
QY	1441	GAACTGTAAAGCTACAGATGGTCTTTACAAATGGAACCCCA	1481			
Db	1970	GAACTGTAAAGCTACAAATGGAGCCCAAGATGAGTCCAA	2010			
RESULT 2						
CR613169						
LOCUS	full-length cDNA clone CS0DE013Y120 of Placenta of Homo sapiens					
DEFINITION	(human).					
ACCESSION	CR613169					
VERSION	1 GI:50493976					
KEYWORDS	HTC; CNSLT cdna.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE						
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.					
TITLE	Full-length cDNA libraries and normalization					
JOURNAL	Unpublished					
REMARK	Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue					
REFERENCE						
AUTHORS	Genoscope.					
TITLE	Direct Submission					
JOURNAL	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr					
COMMENT						
- Web : www.genoscope.cns.fr)						
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.						
FEATURES						
source						
1..2716						
/organism="Homo sapiens"						
/mol_type="mRNA"						
/db_xref="taxon:9606"						
/clone="CS0DE013Y120"						
/tissue_type="Placenta"						
/plasmid="pCMVSPORT_6"						
ORIGIN						
Query Match 90.1%; Score 1333.8; DB 3; Length 2716;						
Best Local Similarity 93.8%; Pred. No. 0;						
Matches 1389; Conservative 0; Mismatches 92; Indels 0; Gaps 0;						
QY	1	ATGCGCCCTCCCTTATCATATCTTTCTTTTACTTTTCTTTTACCTCCCTTTGCTTCACT	60			
Db	773	ATGCGCCCTCCCTTATCATATTTTCTTTTACTTTTCTTTTACCTCTTTTCACTTCACT	832			

```
QY 61 GCACCCCTCCATGCTGCTGACCAACAGTAGTCCCTTACCAAGAGTTTCTATGAGA 120
Db 833 GCACCCCTCCATGCTGCTGACCAACAGTAGTCCCTTACCAAGAGTTTCTATGAGA 892
QY 121 ACAGCGCTTCCCTGGAATATTTGATGCCCCATATATAGGAGTTTATCTAAGGGAATCTCC 180
Db 893 ATGAGCGTCCCGGAATATTTGATGCCCCATATATAGGAGTTTATCTAAGGGAATCTCC 952
QY 181 ACCTTCACTGCCACACCCATATATGCCCCGCAATCTGCTATTAATCTCTGCACTCTTTGGATG 240
Db 953 ACCTTCACTGCCACACCCATATATGCCCCGCAATCTGCTATTAATCTCTGCACTCTTTGGATG 1012
QY 241 CATGCAATATCTATTTATGCAAGGGAATATGATTAATCTCTGCTGCACTCTTTGGATG 300
Db 1013 CATGCAATATCTATTTATGCAAGGGAATATGATTAATCTCTGCTGCACTCTTTGGATG 1072
QY 301 GGAAGCACTCTCTGTTGGATTTACTTACCCCATATACAGTATGCTGATGGGGGTGGAAT 360
Db 1073 GGAAGCACTCTCTGTTGGATTTACTTACCCCATATATGATGCTGATGGGGGTGGAAT 1132
QY 361 CAAGGTGAGCAAGAGAAACAAAGTAAGGAAGCAATCTCCCAACTGACCCGGGACAT 420
Db 1133 CAAGGTGAGCAAGAGAAACAAAGTAAGGAAGCAATCTCCCAACTGACCCGGGACAT 1192
QY 421 AGCACCTAGCCCTACAAAGCACTAGTTCTCTCAAACTACATGAACCTCCGTACC 480
Db 1193 GGCACCTAGCCCTACAAAGCACTAGTTCTCTCAAACTACATGAACCTCCGTACC 1252
QY 481 CATACTCGCTGCTGAGCCTATTTAATACCACTCCTCCTCGGCTCATGAGGTCTCAGCC 540
Db 1253 CATACTCGCTGCTGAGCCTATTTAATACCACTCCTCCTCGGCTCATGAGGTCTCAGCC 1312
QY 541 CAAACCCCTACTAATCTGTTGGATGCTGCTCCCTGCACTTTCAGGCCATACATTTCAATC 600
Db 1313 CAAACCCCTACTAATCTGTTGGATGCTGCTCCCTGCACTTTCAGGCCATACATTTCAATC 1372
QY 601 CTTGTTCTCTGAAACAAAGTAAGCACTTTCAGCAGAAATAAAGCACTTCCGTTTATGTA 660
Db 1373 CTTGTTCTCTGAAACAAAGTAAGCACTTTCAGCAGAAATAAAGCACTTCCGTTTATGTA 1432
QY 661 GGAACCTCTGTTTCCAACTGGAATAAACCCTACCTCCTCAACCTCCTGTTGTAATTT 720
Db 1433 GGAACCTCTGTTTCCAACTGGAATAAACCCTACCTCCTCAACCTCCTGTTGTAATTT 1492
QY 721 AGCAATATCTATAGACAAACCACTCCCAATGATCAGGTGGGTAAACCTCCCAACGA 780
Db 1493 AGCAATATCTATAGACAAACCACTCCCAATGATCAGGTGGGTAAACCTCCCAACGA 1552
QY 781 ATAGTCTGCTACCTCAGGAATATTTTGTCTGCTGCTACCTCAGCCTATCATTTGTTG 840
Db 1553 ATAGTCTGCTACCTCAGGAATATTTTGTCTGCTGCTACCTCAGCCTATCATTTGTTG 1612
QY 841 AATGGCTCTTCAAACTATGCTCTCTCTCAATCTTATGTCCTCCTATGACCATCTAC 900
Db 1613 AATGGCTCTTCAAACTATGCTCTCTCTCAATCTTATGTCCTCCTATGACCATCTAC 1672
QY 901 ACTGAACAAGATTTATACATATCTGCTAATCTTAAGCCCAACCAAAAGAGTACCATT 960
Db 1673 ACTGAACAAGATTTATACATATCTGCTAATCTTAAGCCCAACCAAAAGAGTACCATT 1732
QY 961 CTTCTCTTTGTTATCAGAGCAGAGTGTAGGAGACTAGGTACTGGCATTTGGCAGTATC 1020
Db 1733 CTTCTCTTTGTTATAGAGCAGAGTGTAGGAGACTAGGTACTGGCATTTGGCAGTATC 1792
QY 1021 ACAACCTCTACTCAGTTCTACTCAAACTATCTCAAGAAATAATGTCATGGAACAG 1080
Db 1793 ACAACCTCTACTCAGTTCTACTCAAACTATCTCAAGAAATAATGTCATGGAACAG 1852
QY 1081 GTCACCTGCTCCTGCTGCTTCTGCAAGATCAACTTAACTCCCTAGCAGCAGTATGCTT 1140
Db 1853 GTCGCGCACTCCTGCTGCTTCTGCAAGATCAACTTAACTCCCTAGCAGCAGTATGCTT 1912
QY 1141 CAAAATCGAAGAGCTTTAGACTTTGCTAAACCGCCAAAGAGGGGGAACCTGTTTATTTT 1200
```

```
Db 1913 CAAAATCGAAGAGCTTTAGACTTTGCTAAACGCTGAAAGGGGGAACCTGTTTATTTT 1972
QY 1201 GGAAGAAAGAGCTGTTTATTTATTAATCAATCCAGAAATTTGCTACTGAGAAAGTTAAAGAA 1260
Db 1973 GGGGAAGAATGCTGTTTATTTATTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAA 2032
QY 1261 ATTGAGATCGAATACAAATGTAGAGCAGAGAGCTTCAAAACACCAACCGCTGGGGCTC 1320
Db 2033 ATTGAGATCGAATACAAATGTAGAGCAGAGAGCTTCAAAACACCAACCGCTGGGGCTC 2092
QY 1321 CTCAGCAATGAGTGCCTGGGTTCTCCCTCTCTTAGGACCTCTAGCAGCTCTAAATATTG 1380
Db 2093 CTCAGCAATGAGTGCCTGGGTTCTCCCTCTCTTAGGACCTCTAGCAGCTCTAAATATTG 2152
QY 1381 TTACTCTCTTTGGACCCCTGTATCTTTAACTCTCTTTAAAGTTTGTCTCTCCAGAAAT 1440
Db 2153 CTACTCTCTTTGGACCCCTGTATCTTTAACTCTCTTTAACTCTCTCTCTCCAGAAATC 2212
QY 1441 GAAGCTGTAAGCTACAGATGGTCTTTACAAATGGAACCCCA 1481
Db 2213 GAAGCTGTAAGCTACAAATGGAACCCCAAGATGCAGTCCAA 2253

RESULT 3
CR605851 2748 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CS0D012YJ24 of Placenta of Homo sapiens
DEFINITION (human).
ACCESSION CR605851
VERSION CR605851.1 GI:50486658
KEYWORDS HTC; CNSLT_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2748)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue
2 (bases 1 to 2748)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen
FEATURES
source
1..2748
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D012YJ24"
/tissue_type="Placenta"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 90.1%; Score 1333.8; DB 3; Length 2748;
Best Local Similarity 93.8%; Pred. No. 0;
Matches 1389; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
QY 1 ATGGCCCTCCCTTATCATACTTTTCTTTACTTTTCTTACCTCCCTTTCGCTCTCACT 60
Db 772 ATGGCCCTCCCTTATCATATTTTCTTTACTTTTCTTTTACCTCTTTCCTCACT 831
QY 61 GCACCCCTCCATGCTGCTGACCAACAGTAGTCCCTTACCAAGAGTTTCTATGAGA 120
```

```
Db      832 GCACCCCTCCATGCGCTGTATGACGAGTAGTCCCTTACCAGAGTTTCTATGAGA 891
Qy      121 AC0GGGCTTCTGGAATATGATGCGCCCATATATAGAGTTTATCTAAGGGAATCTCC 180
Db      892 ATGACGGCTCCCGGAAATATGATGCGCCCATATGATAGGAGTCTTTCTAAGGGAACCCCC 951
Qy      181 ACCTTCACTGCCCCACACCCATATGCCCCGCAACTGCTATAAATCTGCACTCTTTGCATG 240
Db      952 ACCTTCACTGCCCCACACCCATATGCCCCGCAACTGCTATAAATCTGCACTCTTTGCATG 1011
Qy      241 CATGCAAAATCTCATTTATGGAAGGGAATAATGATTAATCTTCTAGTTTGTCTGTGAGGACTT 300
Db      1012 CATGCAAAATCTCATTTATGGAAGGGAATAATGATTAATCTTCTAGTTTGTCTGTGAGGACTT 1071
Qy      301 GGAGCCACTGCTGTGGAATCTTACTTACCCTACCATCATGATGCTGATGCGGGTGAAT 360
Db      1072 GGAGTCACTGCTGTGGAATCTTACTTACCCTACCATCATGATGCTGATGCGGGTGAAT 1131
Qy      361 CAAGGTCAGGCAAGGAAACAAAGTAAGGAAGCAATCTCCCAACTGACCCGGGACAT 420
Db      1132 CAAGATCAGGCAAGGAAACAAAGTAAGGAAGCAATCTCCCAACTGACCCGGGACAT 1191
Qy      421 AGCACCCCTAGCCCTACAAAGGACTAGTTCTCTCAAACTACATGAAACCCCTCCGTACC 480
Db      1192 GGCACCTCTAGCCCTACAAAGGACTAGATCTCTCAAACTACATGAAACCCCTCCGTACC 1251
Qy      481 CATACTGCGCTGTGAGCCCTATTTATACACCCCTCACTCGGCTCCATGAGGTCACGCC 540
Db      1252 CATACTGCGCTGTGAGCCCTATTTATACACCCCTCACTCGGCTCCATGAGGTCACGCC 1311
Qy      541 CAACACCCCTACTACTGCTGTGATGCTGCTCCCTGCACCTTCAGGCTCAGCATATTTCAATC 600
Db      1312 CAACACCCCTACTACTGCTGTGATGCTGCTCCCTGCACCTTCAGGCTCAGCATATTTCAATC 1371
Qy      601 CCTGTTCTGCAACATGGAACAACTTCAGCACAGAAATAAACACCACTTCGGTTTGTGTA 660
Db      1372 CCTGTTCTGCAACATGGAACAACTTCAGCACAGAAATAAACACCACTTCGGTTTGTGTA 1431
Qy      661 GGACCTCTGTTTCCATCTGGAATAACCCATACCTCAACCTCAGCTGTTGTAATTT 720
Db      1432 GGACCTCTGTTTCCATCTGGAATAACCCATACCTCAACCTCAGCTGTTGTAATTT 1491
Qy      721 AGCAATCTATAGACACACACGAGCTCCCAATGCAATCAGGTGGGTAAACACCTCCACACA 780
Db      1492 AGCAATCTATACATACACACCACTCCCAATGCAATCAGGTGGGTAAACACCTCCACACA 1551
Qy      781 ATAGTCTGCTTACCTCAGGAATATTTTGTCTGTGTTACCTCAGCCTATCATTTGTTG 840
Db      1552 ATAGTCTGCTTACCTCAGGAATATTTTGTCTGTGTTACCTCAGCCTATCATTTGTTG 1611
Qy      841 AATGCTCTTCAGATCTATGCTGCTCTCTCATTTCTAGTGCCCTTATGACCATCTAC 900
Db      1612 AATGCTCTTCAGATCTATGCTGCTCTCTCATTTCTAGTGCCCTTATGACCATCTAC 1671
Qy      901 ACTGAACAAGATTTATACAATCATGTGTAATCTAGCCGCCCAACAACAAAGAGTACCCATT 960
Db      1672 ACTGAACAAGATTTATACAGTTATGTATATCTAAGCCCGCAACAACAAAGAGTACCCATT 1731
Qy      961 CTTCTCTTTGTTATCAGACGAGGAGTGTCTAGGCGACATGAGTACTGGCATTTGGCAGTATC 1020
Db      1732 CTTCTCTTTGTTATGAGGAGGAGTGTCTAGGCGACATGAGTACTGGCATTTGGCAGTATC 1791
Qy      1021 ACAACCTCTACTCAGTTCTTACTACAACTATCTCAAGAAATAAATGTTGACATCGAACAG 1080
Db      1792 ACAACCTCTACTCAGTTCTTACTACAACTATCTCAAGAAATAAATGTTGACATCGAACAG 1851
Qy      1081 GTCAGTACTCCCTGGTCACTTGTCAAGATCAACTTAATCTCCCTAGCAGCAGTACTGCTT 1140
Db      1852 GTGCGGCACTCCCTGGTCACTTGTCAAGATCAACTTAATCTCCCTAGCAGCAGTACTGCTT 1911
Qy      1141 CAAATTCAGAGCTTTAGATCTGTTAATCCGCAAAAGAGGGGAACTGTTTATTTTA 1200
```

```
Db      1912 CAAAATCGAAGAGCTTTAGACTTGTCTAAACCGCTGAAAGAGGGGAAACCTGTTTATTTTA 1971
Qy      1201 GGAGAGAAACGCTGTTTATTTATCTTAATCAATCCAGAAATGTCTACCTGAGAAAGTAAAGAA 1260
Db      1972 GGGAGAAATGCTGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 2031
Qy      1261 ATTTCAGATTCGAATACAATGTAGAGAGGAGCTTTCAAAACACCAAGACGCTGGGCGCTC 1320
Db      2032 ATTTCAGATTCGAATACAACGTAGAGAGGAGCTTTCAAAACACCAAGACGCTGGGCGCTC 2091
Qy      1321 CTCAGCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
Db      2092 CTCAGCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2151
Qy      1381 TTACTCTCTTTGGAGCCCTGTATCTTTAACTCTCTTTAACTCTCTTTAACTCTCTTTAACTCTCT 1440
Db      2152 CTACTCTCTTTGGAGCCCTGTATCTTTAACTCTCTTTAACTCTCTTTAACTCTCTTTAACTCTCT 2211
Qy      1441 GAAGCTGTTAACTACAGATGCTTCTTACAAATGGAACCCCA 1481
Db      2212 GAAGCTGTTAACTACAAATGGAGCCCAAGATGCAGTCCAA 2252

RESULT 4
CR625046 2748 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CS0D1044YK06 of Placenta Cot 25-normalized
DEFINITION of Homo sapiens (human).
ACCESSION CR625046
VERSION CR625046.1 GI:50505853
KEYWORDS HTC; CNSLT_cDNA.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2748)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue
Genoscope.
2 (bases 1 to 2748)
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source
1..2748
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1044YK06"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 90.1%; Score 1333.8; DB 3; Length 2748;
Best Local Similarity 93.8%; Pred. No. 0;
Matches 1389; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy      1 ATGGCCCTCCCTTATCATCTTTCTTTTACTTGTCTTCTTACCTCTTCTTACCTCTTCTGCTCTCACT 60
Db      783 ATGGCCCTCCCTTATCATCTTTCTTTTACTTGTCTTCTTACCTCTTCTTACCTCTTCTGCTCTCACT 842
Qy      61 GCACCCCTCCATGCTGCTGTACAAACCAAGTAGTCCCTTTACCAAGAGTTTCTATGAGA 120
```

843 GCACCCCTCATGCGCTGTATGACCAAGTCTCCCTTACCAAGAGTTTCTATGAGA 902
121 ACGGGCTTCTGGAATAATGATGCCCATATATAGGAGTTTATCAAGGGAACCTCC 180
903 ATGACGGCTCCGGAATAATGATGCCCATATGATAGGAGTCTTTCTAAGGGAACCCCC 962
181 ACCTTCACTGCCACACCCATATGCCCCGCAACTGCTATTAACCTGCACTCTTTGCAATG 240
963 ACCTTCACTGCCACACCCATATGCCCCGCAACTGCTATTAACCTGCACTCTTTGCAATG 1022
241 CATGCAATACTCATTAATTTGGACAGGGAATAATGATTAATCTAGTTTCTCTGGAGACTT 300
1023 CATGCAATACTCATTAATTTGGACAGGGAATAATGATTAATCTAGTTTCTCTGGAGACTT 1082
301 GGAGCCACTGCTGTTGGACCTTACCTCAACCATACCAAGTATGCTGATGAGGGTGGAAAT 360
1083 GGAGTCACTGCTGTTGGACCTTACCTCAACCATACCAAGTATGCTGATGAGGGTGGAGTT 1142
361 CAAGTCCAGCAAGAGAAACAAAGTAAAGGAAGCAATCTCCCACTGACCCGGGACAT 420
1143 CAAGATCAGCAAGAGAAACAAAGTAAAGGAAGCAATCTCCCACTGACCCGGGACAT 1202
421 AGCACCCCTAGCCCTACAAAGGACTAGTTCTCTCAAAACTACATGAACCCCTCCGTACC 480
1203 GGACACTCTAGCCCTACAAAGGACTAGTCTCTCAAAACTACATGAACCCCTCCGTACC 1262
481 CATACTCGCTGCTGAGCCCTATTATTAATACCAACCTCTCACTCGGCTCCATGAGGTTCTAGCC 540
1263 CATACTCGCTGCTGAGCCCTATTATTAATACCAACCTCTCACTCGGCTCCATGAGGTTCTGGCC 1322
541 CAACACCTACTACTGCTGAGGAGTGGCTCCCTCCCTGCACTTCAGGCCATACATTTCATC 600
1323 CAACACCTACTACTGCTGAGGAGTGGCTCCCTCCCTGCACTTCAGGCCATACATTTCATC 1382
601 CTGTTCTCTCAACAAATGGAACAACTTCAGCACAGAAATAAACCACTCTCCGTTTGTAGTA 660
1383 CTGTTACTGACAAATGGAACAACTTCAGCACAGAAATAAACCACTCTCCGTTTGTAGTA 1442
661 GGACCTCTGTTTCCAACTGGAATAAACCACTACCTCAACCTCACTGCTGTGTAATAATTT 720
1443 GGACCTCTGTTTCCAACTGGAATAAACCACTACCTCAACCTCACTGCTGTGTAATAATTT 1502
721 AGCAATCTATAGACAAACAGCTCCCAATGATCAGGTGGGTAAACCTCCACACGA 780
1503 AGCAATCTATAGACAAACAGCTCCCAATGATCAGGTGGGTAAACCTCCACACGA 1562
781 ATAGTCTGCTTACCTCAGCAATATTTTGTCTGTGTTACCTCAGCTCATCATTTGTTTG 840
1563 ATAGTCTGCTTACCTCAGCAATATTTTGTCTGTGTTACCTCAGCTCATCATTTGTTTG 1622
841 AATGGCTCTTTCAGAACTATGCTGCTCTCTCATTTCTTAGTGCCCTCATGACCATCTAC 900
1623 AATGGCTCTTTCAGAACTATGCTGCTCTCTCATTTCTTAGTGCCCTCATGACCATCTAC 1682
901 ACTGAACAAGATTATACAACTATGCTGTAACCTAAGCCCAACAAAGAGTACCCATT 960
1683 ACTGAACAAGATTATACAACTATGCTGTAACCTAAGCCCAACAAAGAGTACCCATT 1742
961 CTTCCTTTGTTTATCAGACAGAGTGTAGGACAGTACTGATGCTGCACTTGGCAGTATC 1020
1743 CTTCCTTTGTTTATGAGAGCAGAGTGTAGGACAGTACTGATGCTGCACTTGGCAGTATC 1802
1021 ACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGTTGATGCAAGACAG 1080
1803 ACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGTTGATGCAAGACAG 1862
1081 GTCACTGACTCCCTGGTCACTTGCAGAGATCAACTTAACCTCCCTAGCAGCAGTAGTCTTT 1140
1863 GTCCCGGACTCCCTGGTCACTTGCAGAGATCAACTTAACCTCCCTAGCAGCAGTAGTCTTT 1922
1141 CAAATCGAAGAGTTTATGCTTGAACCTGCTTAACCCCAAGAGGGGACCTGTTTATTTTAA 1200
1923 CAAATCGAAGAGTTTATGCTTGAACCTGCTTAACCCCAAGAGGGGACCTGTTTATTTTAA 1982

1201 GGAGAAGAAGCCTGTTTATTATTATGTTAATCAATCAGAAATTTCTACTGAGAAAGTTAAAGAA 1260
1983 GGGAAGAAGTCTGTTTATTATTATTATGTTAATCAATCAGAAATTTCTACTGAGAAAGTTAAAGAA 2042
1261 ATTTCGAGATCGAATACAAATGCTAGACAGAGAGCTTCAAAACACCCGAAACCTCTGGGCTC 1320
2043 ATTTCGAGATCGAATACAAATGCTAGACAGAGAGCTTCAAAACACCCGAAACCTCTGGGCTC 2102
1321 CTGAGCAATGAGTGGCTGGCTTCTCCCTCTCTTAGGACCTCTAGCAGCTCTAATAATTG 1380
2103 CTGAGCAATGAGTGGCTGGCTTCTCCCTCTCTTAGGACCTCTAGCAGCTCTAATAATTG 2162
1381 TTTACTCTCTTTGGACCTGTTATCTTTAACCTCTCTTTAAGTTTGTCTCTTCAGAAATT 1440
2163 CTACTCTCTTTGGACCTGTTATCTTTAACCTCTCTTTAAGTTTGTCTCTTCAGAAATC 2222
1441 GAAGTGTAAAGCTACAGATGGTCTTTACAAATGGAACCCCA 1481
2223 GAAGTGTAAAGCTACAAATGGAACCCCAAGATGAGTCCAA 2263

RESULT 5
CR617248
LOCUS
DEFINITION
full-length cDNA clone CS0D1022YJ18 of Placenta Cot 25-normalized
of Homo sapiens (human)
ACCESSION
CR617248
VERSION
CR617248.1 GI:50498055
KEYWORDS
HTC; CNSLT cDNA
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 2749)
Li W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
AUTHORS
JOURNAL
REMARK
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue
REFERENCE
2 (bases 1 to 2749)
Genoscope.
Direct Submission
TITLE
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
source
1..2749
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1022YJ18"
/issue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 90.1%; Score 1333.8; DB 3; Length 2749;
Best Local Similarity 93.8%; Pred. No. 0;
Matches 1389; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
1 ATGGCCCTCCCTTATCATCTTTTCTCTTTTACTGTTCTCTTACCCCTTTCGCTCTCACT 60
777 ATGGCCCTCCCTTATCATATTTTCTCTTTTACTGTTCTTTTACCCCTTTCGCTCTCACT 836
61 GCACCCCTCCCATGCTGCTGTAACACCACTAGTCTCCCTTACCAGAGTTTCTTATGAGA 120
837 GCACCCCTCCCATGCTGCTGTAACACCACTAGTCTCCCTTACCAGAGTTTCTTATGAGA 896

```
QY 121 ACGGCGCTTCTGGAATATTGATGCGCCCATCATATAGGAGTTTATCTTAAGGGAATCC 180
Db 897 ATGACGGCTCCCGGAATATTGATGCGCCCATCGTATAGGAGTCTTTCTTAAGGGAACCC 956
QY 181 ACCTTCACCTGCCCCACACCCCATATGCCCCGCAACTGCTATAAATCTCTGCCACTCTTTGCA 240
Db 957 ACCTTCACCTGCCCCACACCCCATATGCCCCGCAACTGCTATAAATCTCTGCCACTCTTTGCA 1016
QY 241 CATGCAAAATCTCATTTATTGGAAGGGAATAATGATTAATCTTAGTTGTCTTGAGAGACTT 300
Db 1017 CATGCAAAATCTCATTTATTGGAAGGGAATAATGATTAATCTTAGTTGTCTTGAGAGACTT 1076
QY 301 GAGGCCACTGCTGTTGGACTTACTTACCCATACACAGTATGCTCATGAGGGGTGGAAT 360
Db 1077 GAGGTCACTGCTGTTGGACTTACTTACCCAACTGTTGTAAGTCTCATGAGGGGTGAGTT 1136
QY 361 CAAGGTGAGCAAGAGAGAAAACAAAGTAAAGGAAGCAATCTCCAACTGACCGGGGACAT 420
Db 1137 CAAGTCAAGCAAGAGAGAAAACAAAGTAAAGGAAGTAACTCTCCAACTGACCGGGGTACAT 1196
QY 421 AGCACCCCTAGCCCTTCAAGAGACTAGTTCTCTCAAAACTACATGAACCCCTCCGTACC 480
Db 1197 GGCACCTCTAGCCCTTCAAGAGACTAGATCTCTCAAAACTACATGAACCCCTCCGTACC 1256
QY 481 CATACTGCGCTGTTGAGCCTATTATTAACACACCTCACTCGGCTCCATGAGGTCTCAGCC 540
Db 1257 CATACTGCGCTGTTGAGCCTATTATTAACACACCTCACTCGGCTCCATGAGGTCTCAGCC 1316
QY 541 CAAACCCCTACTACTGTTGGATGTGCTCCCTCCCTGCATCTCAGGCCATACATTTCAATC 600
Db 1317 CAAACCCCTACTACTGTTGGATGTGCTCCCTCCCTGCATCTCAGGCCATATGTTTCAATC 1376
QY 601 CTTGTTCTGGAACATGGAACACTTCAGACAGAAATAACACCACTTCCTGTTTAGTA 660
Db 1377 CTTGTTCTGGAACATGGAACACTTCAGACAGAAATAACACCACTTCCTGTTTAGTA 1436
QY 661 GGACCTCTGTTTCCCAATCTGGAATAACCATACCTCAAACTCTCACTGTGTAATAAT 720
Db 1437 GGACCTCTGTTTCCCAATCTGGAATAACCATACCTCAAACTCTGTAATAAT 1496
QY 721 AGCAATACTATAGACACCAACAGCTCCCAATGATCAGGTGGGTAAACCTCCACACGA 780
Db 1497 AGCAATACTATAGACACCAACAGCTCCCAATGATCAGGTGGGTAACTCTCCACACAA 1556
QY 781 ATAGTCTGCTTACCTCAGGATATTTTGTCTGCTGCTACCTCAGCTCATCTGTTTG 840
Db 1557 ATAGTCTGCTTACCTCAGGATATTTTGTCTGCTGCTACCTCAGCTCATCTGTTTG 1616
QY 841 AATGGCTCTTACAGATCTATGCTTCTCTCTCATCTTAGTGCCCTCTATGACCATCTAC 900
Db 1617 AATGGCTCTTACAGATCTATGCTTCTCTCTCATCTTAGTGCCCTCTATGACCATCTAC 1676
QY 901 ACTGAACAAGATTTATACAACTATGCTGATCTTAAGCCCAACAAAGAGTACCCATT 960
Db 1677 ACTGAACAAGATTTATACAGTATGCTATATCTTAAGCCCAACAAAGAGTACCCATT 1736
QY 961 CTTCTCTTTGTTATCAGAGAGAGGTGCTAGCAGACTAGTACTGTCATTGGCAGTATC 1020
Db 1737 CTTCTCTTTGTTATGAGGAGAGAGGTGCTAGTGCATGAGTACTGTCATTGGCGGTATC 1796
QY 1021 ACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGGTGAATGGAACAG 1080
Db 1797 ACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGGGACATGGAACGG 1856
QY 1081 GTCAGTACTCCCTGGTCACTGTCAGAGATCACTTAATCTCTCTAGCAGGAGTACTCTT 1140
Db 1857 GTGCGGACTCCCTGGTCACTGTCAGAGATCACTTAATCTCTCTAGCAGGAGTACTCTT 1916
QY 1141 CAATATCGAAGCTTTAGACTTGTCTAACCCGCAAGAGGGGAACTGTTTATTTTA 1200
Db 1917 CAATATCGAAGCTTTAGACTTGTCTAACCCGCTGAAGAGGGGAACTGTTTATTTTA 1976
```

```
QY 1201 GGAGAGAAGAACGCTGTTATTATGTTAATCAATCCAGATTTGCTGAGAAAAGTTAAAGAA 1260
Db 1977 GGGGAAGAATGCTGTTATTATGTTAATCAATCCGGAATCGTCACTGAGAAAAGTTAAAGAA 2036
QY 1261 ATTCCAGATCGAATACATATGATAGACAGAGGAGCTTCAAAACACCAAGACGCTGGGGCTC 1320
Db 2037 ATTCCAGATCGAATACAAACGATAGACAGAGGAGCTTCAAAACACCTGGACCCCTGGGGCTC 2096
QY 1321 CTCAGCCAAATGATGCCCTGGGTTCTCCCTTTCTTAGGACCTCTAGCAGCTTAATATTG 1380
Db 2097 CTCAGCCAAATGATGCCCTGGATTTCTCCCTTTCTTAGGACCTCTAGCAGCTTAATATTG 2156
QY 1381 TTACTCTCTTTGGACCCCTGATATCTTTAACTCTCTTTGTTAGTTTGTCTCTCCAGAATT 1440
Db 2157 CTACTCTCTTTGGACCCCTGATATCTTTAACTCTCTTTGTTAACTCTTTGTCTCTCTCCAGAATC 2216
QY 1441 GAAGCTGTAAGCTACAGATGCTTCTTACAAATGGAACCCCA 1481
Db 2217 GAAGCTGTAAGCTACAAATGGAGCCCAAGATGAGTCCAA 2257

RESULT 6
BX391741/c
LOCUS
DEFINITION
BX391741 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1051YM13 3-PRIME, mRNA sequence.
ACCESSION
BX391741
VERSION
BX391741.1 GI:30619529
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 935)
AUTHORS
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished (2001)
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by life technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
4215.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?8=CS0BA1022ZA10_CS02089_1&c=4215.r

FEATURES
source
Location/Qualifiers
1..935
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1051YM13"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 56.7%; Score 839; DB 5; Length 935;
Best Local Similarity 94.2%; Pred. No. 5.6e-237;
Matches 880; Conservative 0; Mismatches 53; Indels 1; Gaps 1;
225 TGCCACTCTTTGATGATGCAATCAATATCTCATTTATGACAGGGAATGATTAATCTCTAG 284
933 TGCCCTCTTTTGGATGATGCAATATCTCA-TATTGGACAGAANNAATGATTAATCTCTAG 875
285 TTGTCTCGGAGGACTTGGAGCCACTGTCTGTGGACTTACTTCTACCCCATACCATGATGTC 344
```

```
Db 874 TTGTCCTGNAGACTTGAAGTCACTGCTGTGTGGACTTACTTACCCAAACTGGTATGTC 815
Qy 345 TGATGGGGTGAATTCAGAGTCAGGCAAGAGAAAAACAAGTAAAGGAACAATCTCTCCCA 404
Db 814 TGATGGGGTGAATTCAGAGTCAGGCAAGAGAAAAACAAGTAAAGGAACAATCTCTCCCA 755
Qy 405 ACTGACCCGGGACATAGACCCCTAGCCCTCAAAAGGACTAGTCTCTCAAAACTACA 464
Db 754 ACTCACCCGGGTACATGSCACCTCTAGCCCTCAAAAGGACTAGATCTCTCAAAACTACA 695
Qy 465 TGAACCCCTCCGTACCCATACCTCGCTGGTGGAGCCTATTAAATACCAACCCCTCACTGGCT 524
Db 694 TGAACCCCTCCGTACCCATACCTCGCTGGTGGAGCCTATTAAATACCAACCCCTCACTGGCT 635
Qy 525 CCATGAGGTCAGCCCAAAACCCCTACTAACTGTGTGGATGTCCTCCCTCGCACTTCAG 584
Db 634 CCATGAGGTCAGCCCAAAACCCCTACTAACTGTGTGGATGTCCTCCCTCGCACTTCAG 575
Qy 585 GCCATACATTTCAATCCCTGTTCTGAAACAATGGAACAATCTCAGACAGAAATAAACAC 644
Db 574 GCCATATGTTTCAATCCCTGTTCTGAAACAATGGAACAATCTCAGACAGAAATAAACAC 515
Qy 645 CACTTCGTTTTAGTAGGACCTCTGTTTCCCAATCTGGAATAACCCATACCTCAACCT 704
Db 514 CACTTCGTTTTAGTAGGACCTCTGTTTCCCAATCTGGAATAACCCATACCTCAACCT 455
Qy 705 CACTCTGTAAAAATTTAGCAATACTATAGACACAACAGCTCCCAATGCAATGCAATGAGTGGGT 764
Db 454 CACTCTGTAAAAATTTAGCAATACTATAGACACAACAGCTCCCAATGCAATGAGTGGGT 395
Qy 765 AACACTCCACACGAAATAGTCTGCTCAACCTCAGGAATATTTTTGTCTGTGTAACCTC 824
Db 394 AACCTCTCCACACAAATAGTCTGCTCAACCTCAGGAATATTTTTGTCTGTGTAACCTC 335
Qy 825 AGCTATCATGTTTGAATGGCTCTTCAGAACTATGTCCTTCTCTCATCTTAGTGCC 884
Db 334 AGCTATCATGTTTGAATGGCTCTTCAGAACTATGTCCTTCTCTCATCTTAGTGCC 275
Qy 885 CCCTATGACCATCTACACTGAAACAAGATTTATACAACTATGTCGTACCTTAAGCCCAACA 944
Db 274 CCCTATGACCATCTACACTGAAACAAGATTTATACAGTATATGTCATATCTAAGCCCGCAA 215
Qy 945 CAAAAGAGTACCATCTCTCTCTTTGTTATCAGACGAGGAGTCTAGGACAGACTAGGTAC 1004
Db 214 CAAAAGAGTACCATCTCTCTCTTTGTTATAGGACGAGGAGTCTAGGTGCTAGGTAC 155
Qy 1005 TGGCATTGGCAGTATCACAACTCTACTCAGTTCTACTACAAACTATCTCAAGAAATAA 1064
Db 154 TGGCATTGGCGGTATCACAACTCTACTCAGTTCTACTACAAACTATCTCAAGAACTAA 95
Qy 1065 TGGTGATGGAACAGTCACTGACTCCCTGGTCACTCTGCAAGATCAACTTAACCTCCCT 1124
Db 94 TGGGGACATGGAACGGGTGCGGACCTCCCTGGTCACTCTGCAAGATCAACTTAACCTCCCT 35
Qy 1125 AGCAGAGTAGTCTCTTCAAAATCGAAGAGCTTTA 1158
Db 34 AGCAGAGTAGTCTCTTCAAAATCGAAGAGCTTTA 1
```

RESULT 7
BX347952

LOCUS BX347952 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS0D1051YML13 5-PRIME, mRNA sequence.

ACCESSION BX347952

VERSION BX347952.1

KEYWORDS GI:30375235

SOURCE EST.

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 955)

AUTHORS
TITLE
JOURNAL
COMMENT

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
4215.r

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?s=CS0BAE0232F07_AE02167_l&c=4215.r

FEATURES
source

Location/Qualifiers

1..955

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0D1051YML13"

/tissue_type="PLACENTA COT 25-NORMALIZED"

/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/notes="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 54.9%; Score 813.6; DB 5; Length 955;

Best Local Similarity 93.7%; Pred. No. 1.9e-229;

Matches 890; Conservative 0; Mismatches 56; Indels 4; Gaps 4;

Qy 226 GGCACCTCTTGCATGTCATGCNAATACCTCATTTATGACAGGGAATGATTAATCTCTAGT 285

Db 1 GCACCTCTTGCATGTCATGCNAATACCTCATTTATGACAGGGAATGATTAATCTCTAGT 60

Qy 286 TGTCTCGGAGGACTTCGAGCCACTGCTGTGTGACCTTACTTCAACCATACAGTATGTCT 345

Db 61 TGTCTCGGAGGACTTCGAGCTACTGCTGTGTGACCTTACTTCAACCATACAGTATGTCT 120

Qy 346 GATGGGGTGGAAATCAAGGTGAGCAAGAGAAAAACAAGTAAAGGAACAATCTCCCA 405

Db 121 GATGGGGTGGAGTTCAAGATCAGGCAAGAAAAACAATGTAAGAGTAATCTCCCA 180

Qy 406 CTGACCCGGGACATAGCACCCCTAGCCCTACAAAGGACTAGTTCTCTCAAAACTACAT 465

Db 181 CTCACCCGGGTACATGGCACCTCTAGCCCTCTACAAAGGACTAGATCTCTCAAAACTACAT 240

Qy 466 GAAACCTCTCGTACCCATCTCGCCTGGTGGAGCTATTTAATACCACTCCTCACTCGGCTC 525

Db 241 GAAACCTCTCGTACCCATCTCGCCTGGTGGAGCTATTTAATACCACTCCTCACTCGGCTC 300

Qy 526 CATGAGGTCTCAGCCCAAAACCTACTAATGTTGGATGTGCTCCCTCCCTGCACTTCAGG 585

Db 301 CATGAGGTCTCGGCCCAAAACCTACTAATGTTGGATGTGCTCCCTCCCTGCACTTCAGG 360

Qy 586 CCATACATTTCAATCCCTGTTCTGACCAATGGAACAATCTCAGCAGAGAAATAAACACC 645

Db 361 CCATATGTTTCAATCCCTGTTCTGACCAATGGAACAATCTCAGCAGAGAAATAAACACC 420

Qy 646 ACTTCGGTTTTAGTAGGACCTCTGTTTCCCAATCTGGAATAACCCATACCTCAACCTC 705

Db 421 ACTTCGGTTTTAGTAGGACCTCTGTTTCCCAATCTGGAATAACCCATACCTCAACCTC 480

Qy 706 ACCTGTGTAAAAATTTAGCAATACTATAGACACAACACGAGTCCCAATGCACTCAGGTGGGTA 765

Db 481 ACCTGTGTAAAAATTTAGCAATACTATAGACACAACACGAGTCCCAATGCACTCAGGTGGGTA 540

Qy 766 ACACCTCCACAGGAATAGTCTGCCTACCCCTCAGGAATATTTTTGTCTGTGTGTAACCTCA 825

Db 541 ACTCCT-CCAAACAAATAGTCTGCCTACCCCTCAGGAATATTTTTGTCTGTGTGTAACCTCA 599

```
QY      826 GCCTATCAITGTTTGAATGCTCTTCAGAACTCTATGCTTCTCTCTCATTTCTTAGTGCCC 885
Db      600 GCCTATCGTGTGTTGAATGCTCTTCAGAACTCTATGCTTCTCTCATTTCTTAGTGCCC 659
QY      886 CCTATGACCATCTACACTGAACTGAAGATTTATACAACTATGCTGCTAGAGCCCAAC 945
Db      660 CCTATGACCATCTACACTGAACTGAAGATTTATACAACTATGCTGCTAGAGCCCAAC 719
QY      946 AAAAGAGTACCCATTCTCTCTTTGTTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACT 1005
Db      720 AAAAGAGTACCCATTCTCTCTTTGTTATGAGGAGCAGGAGTGCTAGGTGCACTAGGTACT 779
QY      1006 GGCATTGGCGGATACCAACTCTACTCTAGTCTTACTTACAACTATCTCAAGAAATAAAT 1065
Db      780 GGCATTGGCGGATACCAACTCTACTCTAGTCTTACTACANACTATCTCAAGAACTAAAT 839
QY      1066 GGTGACAT-GGAACAGGTCACTGACTCCCTGG-TCACCTTGGCAAGATCAACTTAATCC 1123
Db      840 GGGGACATGGGAACGGTTCCCGACTTCTCTGGTTCACTTGGCAAGATCAACTTAACCTCC 899
QY      1124 TAGCAGCAGTAG-TCCCTCAAAATCGAAGAGCTTTAGACTTGCTAAACCGC 1172
Db      900 TAGCAGCAGNAGTTCCTTCAAAATCGAAGAGGCTTAGACTTGCTACCCGC 949
```

RESULT 8
BX389656
LOCUS
DEFINITION
BX389656 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1051YM13 5-PRIME, mRNA sequence.

ACCESSION
BX389656
VERSION
BX389656.1 GI:30462930
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 931)
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
4215.r

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0BAF0272B05_AF02531_l&c=4215.r

FEATURES
source
Location/Qualifiers
1. .931
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1051YM13"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 50.8%; Score 752.2; DB 5; Length 931;
Best Local Similarity 92.3%; Pred. No. 3.1e-211;
Matches 856; Conservative 0; Mismatches 65; Indels 6; Gaps 6;

```
QY      267 GAAAATGATTAATCCTAGTTGTTCTGGAGCACTTGGAGCCACTGCTCTGTTGGACTTACTTT 326
Db      1 GAAAAGATTAATCCTAGTTGTTCTGGAGCACTTGGAGCACTGCTCTGTTGGACTTACTTT 60
QY      327 CACCCATACAGTATGCTGATGGGGTGAATTCAGGTGTCAGGCAAGAGAAAAACAAGT 386
Db      61 CACCCAAACTGATGATGCTGATGGGGTGAAGTTCAGGATCAGGCAAGAAAAACAATGT 120
QY      387 AAAGGAAGCAATCTCCCAACTGACCCGGGACATAGCACCCCTAGCCCTTACAAAGACT 446
Db      121 AAAGGAAGTAATCTCCCAACTCACCAGGCTCATGGCACCTCTAGCCCTTACAAAGACT 180
QY      447 AGTTCTCTCAAACTACATGAACCCCTCGTACCCTATCTCGCTCGGTGAGCCCTATTAA 506
Db      181 AGATCTCTCAAACTACATGAACCCCTCGTACCCTATCTCGCTCGGTGAGCCCTATTAA 240
QY      507 TACCAACCTCAGTCCGCTCCATGAGGTCTCAGCCCAAAACCCCTACTTAACTGTTGGATGTG 566
Db      241 TACCAACCTCAGTCCGCTCCATGAGGTCTCGGCCCAAAACCCCTACTTAACTGTTGGATGTG 300
QY      567 CCTCCCTCTGCACTTTCAGGCCATATCAATTTCAATCCCTGTTCTTGAAACAATGGAAACAATT 626
Db      301 CCTCCCTCTGCACTTTCAGGCCATATGTTTCAATCCCTGTTCTTGAAACAATGGAAACAATT 360
QY      627 CAGCAGAGAAATAACACCACTTCCGTTTGTAGTAGGACCTCTTGTTCCTCAATCTGGAAT 686
Db      361 CAGCAGAGAAATAACACCACTTCCGTTTGTAGTAGGACCTCTTGTTCCTCAATCTGGAAT 420
QY      687 AACCCATCTCAAACTCAGCTCTGTTGTAATTTAGCAATCTATATAGACACCAACAGCTC 746
Db      421 AACCCATCTCAAACTCAGCTCTGTTGTAATTTAGCAATCTATATAGACACCAACAGCTC 480
QY      747 CCAATGCATCAGTGGGTAAACACCTCCCAACAAGATGCTGCTTACCCTCAGGAATATT 806
Db      481 CCAATGCATCAGTGGGTAACTCTCCCAACAAGATGCTGCTTACCCTCAGGAATATT 540
QY      807 TTTTGTCTGTGTTACCTCAGCCCTATCAATTTGTTGAATGGCTCTTCAGAACTCTATGCTT 866
Db      541 TTTTGTCTGTGTTACCTCAGCCCTATCGTTGTTGAAATGGCTCTTCAGAACTCTATGCTT 600
QY      867 CTTCTCATTTCTTAGTCCCTCTATGACCATCTACACTGAAACAAGATTTATACAACTATGT 926
Db      601 CTTCTCATTTCTTAGTCCCTCTATGACCATCTACACTGAAACAAGATTTATACAACTATGT 660
QY      927 CGTACCTAAGCCCCACAAACAAAGAGTACCCTTCTCTTTTGTGTTATCAGAGCAGAGT 986
Db      661 CATATCTAAGCCCCGCAACAAAGAGTACCCTTCTCTTTTGTGTTATAGGAGCAGAGT 720
QY      987 GCTAGGACAGTACTAGTACTGGCATTGGCAGTATCAACCTCTACTCAGTTCTTACTACAA 1046
Db      721 GCTAGGTCACACTAGTACTGGCAATTGGC-GGATCACAACCTCTACTCAGTTCTTACTACAN 779
QY      1047 ACTATCTCAAGAAATAAATGGTGCAC-ATGGAAACAGGTCACTGACT-CCCTGGTCACCTTG 1104
Db      780 ACTATCTCAAGAAATAAATGGGAGCAATGTAAACGGTCCGCACTCCCTGGTCACTTC 839
QY      1105 CAAAGATCAACTTAAC-TCCTTAGCAGCAGTA-CTCCTTCAAAATCGAAG-AGCTTTAGAC 1161
Db      840 GCAGATCAACTTACCTTCCCTAGCAGCAGGAGTCTCTTCAAAATCGAAGAAAGCTTTAGCC 899
QY      1162 TTGCTTAACCCCAAAAGAGGGGAACC 1188
Db      900 TTGCTTAACCCGCTGCAAGAGGGGGAAC 926
```

RESULT 9
BX326647/c
LOCUS
DEFINITION
BX326647 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1051YM13 3-PRIME, mRNA sequence.
VERSION
BX326647
KEYWORDS
EST.

```
SOURCE      Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 872)
AUTHORS     Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAD010ZC11_AD00948_1&c=4215.r

FEATURES
  source      Location/Qualifiers
            1..872
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0DI051YM13"
            /tissue_type="PLACENTA COT 25-NORMALIZED"
            /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
            /notes="1st strand cDNA was primed with a NotI-oligo(dT)
            primer. Five prime end enriched, double-strand cDNA was
            digested with Not I and cloned into the Not I and EcoRV
            sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
  Query Match      50.5%; Score 747.2; DB 5; Length 872;
  Best Local Similarity 92.3%; Pred. No. 9.3e-210;
  Matches 808; Conservative 0; Mismatches 64; Indels 3; Gaps 2;

QY 278 ATCTAGTCTCTCGGAGGACTTGGAGCCACTGTCTTGGACTTACTTCCACCACCA 337
    |||||
DB 872 ATCTAGTCTCTCGAGGACTTGAAGTCACTGTCTTGGACTTACTTCCACCACCA 813

QY 338 GTATGTCGTATGGGGTGGAAATCAAGGTGAGGCAAGAGAAAAACAAGTAAGAGACAA 397
    |||||
DB 812 GGTATGGCTATGGGGTGGAGTTCAAGATCAGGCAAGAAAAACAATGTAAA-GAAGTAA 754

QY 398 TCTCCCAACTGACCCGGGACATAGCACCCCTAGCCCTACAAGGACTAGTTCTCTCAA 457
    |||||
DB 753 TCTCCCAACTCACCCGGGTACATGGCACCTTAGCCCTTACAAGGACTAGATCTCTCAN 694

QY 458 AACTACATGAACCCCTCCGTACCCATACCTCGCTGGTGGCTATTTAATACCACCCCTCA 517
    |||||
DB 693 AACTACATGAACCCCTCCGTACCCATACCTCGCTGGTGGCTATTTAATACCACCCCTCA 634

QY 518 CTCGGCTCCATGAGGTCTCAGCCCAAAACCCCTACTAACTGTGGATGTGCTCCCCCTGC 577
    |||||
DB 633 CTGGGCTCCATGAGGTCTCGGCCCAAAACCCCTACTAACTGTGGATGTGCTCCCCCTGA 574

QY 578 ACTTCAGGCCATACATTTCAATCCCTGTTCTGACAAATGGAACCACTTCAGCACAGAAA 637
    |||||
DB 573 ACTTCAGGCCATATGTTTCAATCCCTGTACCTGAACCAATGGAACCACTTCAGCACAGAAA 514

QY 638 TAAACACCACTTCGGTTTTAGTAGGACTCTTGTGTTTCCAATCTGGAATAAACCATACCT 697
    |||||
DB 513 TAAACACCACTTCGGTTTTAGTAGGACTCTTGTGTTTCCAATCTGGAATAAACCATACCT 454

QY 698 CAAACCTCACCTGTGTAAAAATTTAGCAATACTATAGACAAACAGCTCCCAATGCATCA 757
    |||||
DB 453 CAAACCTCACCTGTGTAAAAATTTAGCAATACTATACACAAACCACTCCCAATGCATCA 394

QY 758 GGTGGGTAAACCTCCACAGCAATAGTCTGCTTACCCTCAGGAAATATTTTGTCTGTG 817
    |||||
```

```
393 GGTGGGTAACTCTCCACACAATAAGTCTGCTACCTCAGGAATATTTTGTCTGTG 334
818 GTACCTCAGCCTATCATTTGTTGAATGGCTCTTCAGAAATCATATGCTCTCATCT 877
333 GTACCTCAGCCTATCATTTGTTGAATGGCTCTTCAGAAATCATATGCTCTCATCT 274
878 TAGTGCCCTCATGACCATCTACACATGAACAAGATTTATACAATCATGTCGTACCTAAGC 937
273 TAGTGCCCTCATGACCATCTACACATGAACAAGATTTATACAATCATATGCTCTAAGC 214
938 CCCAACAAAAAGTAGTACCCATTTCTCTTTTGTATTCAGAGCAGGAGTCTTAGGCAGAC 997
213 CCCGCAACAAAAAGTAGTACCCATTTCTCTTTTGTATAGGAGCAGGAGTCTTAGTGCAC 154
998 TAGGTACTGGCANTGGCAGTATCACAACTCTACTAGTCTTACTATAAAATCTCTCAAG 1057
153 TAGGTACTGGCANTGGCAGTATCACAACTCTACTAGTCTTACTATAAAATCTCTCAAG 94
1058 AATAAATGTTGATGACATGGAACAGTCACTGACTCCCTGGTCACTTGCAGAGATCAACTTA 1117
93 AACTAAATGGGACATGGAACGGTGGCGGCTGCGGACTCC--TGGTCACCTGCAAGATCAACTTA 36
1118 ACTCCCTAGCAGCAGTAGTCTCTTCAAAAATCGAAGA 1152
35 ACTCCCTAGCAGCAGTAGTCTCTTCAAAAATCGAAGA 1

RESULT 10
BX408752 903 bp mRNA linear EST 01-MAY-2004
LOCUS    BX408752 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE013Y120
DEFINITION
5-PRIMS, mRNA sequence.
ACCESSION
VERSION  BX408752.2 GI:46931100
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 903)
AUTHORS   Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   On May 13, 2003 this sequence version replaced gi:30648825.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAA008ZH03_CS00697_2&c=4215.r

FEATURES
  source      Location/Qualifiers
            1..903
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0DE013Y120"
            /tissue_type="PLACENTA"
            /clone_lib="Homo sapiens PLACENTA"
            /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
            with a NotI-oligo(dT) primer. Five prime end enriched,
            double-strand cDNA was digested with Not I and cloned into
            the Not I and EcoRV sites of the pCMVSPORT 6 vector.
            Library was not normalized."

ORIGIN
  Query Match      50.0%; Score 741; DB 5; Length 903;
```

Best Local Similarity 93.0%; Pred. No. 6.5e-208; Matches 796; Conservative 0; Mismatches 58; Indels 2; Gaps 2;	
Qy	4 GCCTCCCTTATCATATTTCTTTCTTTACTGTTCTTACCCCTTTCCTCTCTACTGCA 63
Db	50 GGCCCTCCCTANATATTTCTTTACTGTTCTTTTACCCCTTTTCACTCTCACTGCA 109
Qy	64 CCCCTCCATGCTGTGTAACAACAGTAGTCCCTTACCAAGAGTTTCTATGAGAAGC 123
Db	110 CCCCTCCATGCGCTGTATGACAGTAGTCCCTTACCAAGAGTTTCTATGAGAAGT 169
Qy	124 CGCTTCTGGAATATGATGCCCATATATAGGAGTTTACTAAGGAAATCTCCAC 183
Db	170 CAGGTCCCGAATATGATGCCCATATGATAGGAGTTTCTAAGGAAATCTCCAC 229
Qy	184 TTCACTGCCACACCCATATGCCCCGCAACTGTATAACTCTGCCACTCTTTCATGAT 243
Db	230 TTCACTGCCACACCCATATGCCCCGCAACTGTATAACTCTGCCACTCTTTCATGAT 289
Qy	244 GCAATACTATTTGGAAGGAAATGATTAATCTAGTTGCTCTGGAGGACTTGA 303
Db	290 GCAATACTATTTGGAAGGAAATGATTAATCTAGTTGCTCTGGAGGACTTGA 349
Qy	304 GCACTGTCTGTGGAATTTACTTCAACCCATACAGTAGTGTCTGATGGGGTGAATCAA 363
Db	350 GTCACTGTCTGTGGAATTTACTTCAACCCATACAGTAGTGTCTGATGGGGTGAATCAA 409
Qy	364 GGTGAGCAAGAGAAACAAAGTAAGAGAGTATCCCACTGACCCGGGACATAGC 423
Db	410 GATCAGGCAAGAGAAACAAATGTAAGAGAGTATCCCACTGACCCGGGACATAGC 469
Qy	424 ACCCTAGCCCTACAGGAGTATGTTCTCTCAAACTACATGAAACCTCCGTACCCAT 483
Db	470 ACCCTAGCCCTACAGGAGTATGTTCTCTCAAACTACATGAAACCTCCGTACCCAT 529
Qy	484 ACTGCTGTGTGAGCTATTTAATACCACTCTCTCGGTCCATGAGGTCTAGCCCAA 543
Db	530 ACTGCTGTGTGAGCTATTTAATACCACTCTCTCGGTCCATGAGGTCTAGCCCAA 589
Qy	544 AACCTACTAATGTGTGATGCTCCCTGCTGACCTTCAAGGCAATATTCATCCCT 603
Db	590 AACCTACTAATGTGTGATGCTCCCTGCTGACCTTCAAGGCAATATTCATCCCT 649
Qy	604 GTTCTGAAATGAAACAACTTCAAGCAAGAAATAAACACACTTCGTTTGTAGTAGGA 663
Db	650 GTTCTGAAATGAAACAACTTCAAGCAAGAAATAAACACACTTCGTTTGTAGTAGGA 709
Qy	664 CCTCTGTTCATCTGGAATAACCATACCTCAAACTCACTGTTGTAATTTAGC 723
Db	710 CCTCTGTTCATCTGGAATAACCATACCTCAAACTCACTGTTGTAATTTAGC 769
Qy	724 AATACATAGACACACAGCTCCCAATGCAATCAGTGGGTAAACACCTCCACACGAATA 783
Db	770 AATACATAGACACACAGCTCCCAATGCAATCAGTGGGTAAACACCTCCACACGAATA 829
Qy	784 GTCTGCTACCTCAGGAATATTTTGTGTGTGTGATCTCAGCTATCATGTTTGAAT 843
Db	830 GTCTGCTACCTCAGG-ATAATTTTGTGTG-GGTACCTCAGCTATTCGTGTTGGAT 887
Qy	844 GGCTCTTCAGAAATCTA 859
Db	888 GGCTCTTCAGAAATCTA 903
RESULT 11	
BX408733	
LOCUS	
DEFINITION BX408733 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE012U24	
ACCESSION BX408733	
VERSION BX408733.1	
KEYWORDS EST. GI:30635957	
SOURCE Homo sapiens (human)	
ORGANISM Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
1 (bases 1 to 883)	
Li,W.B., Gruber,C., Jesse,J. and Polayes,D.	
Full-length cDNA libraries and normalization	
Unpublished (2001)	
Contact: Genoscope	
Genoscope - Centre National de Sequencage	
BP 191 91006 EVRY cedex - France	
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr	
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime	
end enriched, double-strand cDNA was digested with NotI and cloned	
into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library	
was not normalized. Library was constructed by Life Technologies, a	
division of Invitrogen.	
This sequence belongs to sequence cluster 4215.r	
For more information about this cluster, see	
http://www.genoscope.cns.fr/cdna?&CS0BAA008ZB02_CS00686_1&c=4215.r	
Location/Qualifiers	
1..883	
/organism="Homo sapiens"	
/mol_type="mRNA"	
/db_xref="taxon:9606"	
/clone="CS0DE012U24"	
/tissue_type="PLACENTA"	
/clone_lib="Homo sapiens PLACENTA"	
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed	
with a NotI-oligo(dT) primer. Five prime end enriched,	
double-strand cDNA was digested with NotI and cloned into	
the NotI and EcoRV sites of the pCMVSPORT 6 vector.	
Library was not normalized."	
ORIGIN	
Query Match 49.8%; Score 737.2; DB 5; Length 883;	
Best Local Similarity 93.2%; Pred. No. 8.7e-207;	
Matches 825; Conservative 0; Mismatches 54; Indels 6; Gaps 5;	
Qy	29 TTACTGTTCTTTACCCCTTTTGGCTTCTCATGACCCCTTCATGCTGCTGATGACCA 88
Db	1 TTACTGTTCTTTTACCCCTTTT--ACTCTACTGCAACCCCTCCATGCCCTGTATGACCA 58
Qy	89 GTAGCTCCCTTACCAAGAGTTTCTATGAGAAACGGGCTTCTGGAATAATTGATGCC 148
Db	59 GTAGCTCCCTTACCAAGAGTTTCTATGAGAAATGAGGCTCCGGAAATATTGATGCC 118
Qy	149 CATCATATAGGAGTTTATCTAAGGGAACTCCACCTTCACTGCCACACCCATATGCC 208
Db	119 CATCGTATAGGAGTTTCTTAAGGGAAACCCCACTTCACTGCCACACCCATATGCC 178
Qy	209 GCAACTGCTATTAATCTGCCACTCTTTGATGATGCAATCTATTATTGGACAGGA 268
Db	179 GCAACTGCTATCACTGCCACTCTTTGATGATGCAATCTATTATTGGACAGGA 238
Qy	269 AATGATTAATCTAGTTGCTTGGAGGACTTGGAGCCTGCTGCTGAGCTTACTTCA 328
Db	239 AATGATTAATCTAGTTGCTTGGAGGACTTGGAGCCTGCTGCTGAGCTTACTTCA 298
Qy	329 CCCATACAGTATGTCTGATGGGGTGGAAATTCAGGTGAGCAAGAGAAACAGATA 388
Db	299 CCCAACTGGTATGTCTGATGGGGTGGAGTTTCAAGATCAGGCAAGAGAAACATGTA 358
Qy	389 AGGAGCAATCTCCCACTGACCCGGGACATAGCCTTACCCCTAGCCCTACAAAGGACTAG 448
Db	359 AAGAAGTAATCTCCCACTGACCCGGGTATGTCACCTTAGCCCTACAAAGGACTAG 418
Qy	449 TTCTCTCAAACTACATGAAACCTTCCGTTACCCATCTCCGCTGGTGGCTATTATTA 508
Db	419 ATCTCTCAAACTACATGAAACCTTCCGTTACCCATCTCCGCTGGTAAAGCTATTAA 478
Qy	509 CCACCTCTACCTCGGCTCCATGAGTCTCAGCCCAACCCCTACTACTGTTGGATGTC 568

Db 479 CCACCTCACTGGGCTCCATGAGGTCTCGGCCCAAAACCCCTACTAACTGTTGGATATGCC 538
Qy 569 TCCCCCTGCATTCAGGCCCATCATTTCAATCCCTGTTCTGGAACAATGGAACAACATTC 628
Db 539 TCCCCCTGAACTTCAGGCCCATATGTTTCAATCCCTGTTACCTGAACAATGGAACAACATTC 598
Qy 629 GCACAGAAATAAACACACACCTTCCTGTTTATAGTAGGACCTCTGTTTCCAAATCTGGAATAA 688
Db 599 GCACAGAAATAAACACACACCTTCCTGTTTATAGTAGGACCTCTGTTTCCAAATCTGGAATAA 658
Qy 689 CCATACCTCAAAACCTCACCTGTTGTAATAATTTAGCAATATCTATAGACAAACACAGCTCCC 748
Db 659 CCATACCTCAAAACCTCACCTGTTGTAATAATTTAGCAATATCTATAGACAAACACAGCTCCC 718
Qy 749 AATGCATCA-GTGGGTAAACCTCCGACACGAAATAGTCTGCTACCTCAGCAATATTT 807
Db 719 AATGCATCAGGGTGGGTAACTCCCTCCACAAATAGTCTGCTACCTCA-GAATATTT 777
Qy 808 TTGTCTGTGGTACCTCAGCTATCATTTGTAATGGCTC-TTCAGAACTATGCTT 866
Db 778 TTGTCTGTGGTACCTCAGCTATCATTTGTAATGGCTC-TTCAGAACTATGCTT 837
Qy 867 CCTCTCATTTTCTAGTGGCCCC-TATGACCATCTACACCTGAACAAG 910
Db 838 CCTCTCATTTTCTAGTGGCCCCCTATGACATCTTACACTGACAG 882

RESULT 12
BX347619/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BX347619 877 bp mRNA linear EST 23-APR-2004
BX347619 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1022YJ18 3-PRIME, mRNA sequence.
BX347619
EST.
Homo sapiens (human)
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with NotI and cloned
into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
4215.r

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0BAD009ZB06_AD00811_1&c=4215.r

FEATURES
source

Location/Qualifiers
1. .877
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1022YJ18"
/tissue type="PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with NotI and EcoRV sites of the pCMVSPORT 6 vector. Library
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 49.2%; Score 728.4; DB 5; Length 877;
Best Local Similarity 93.2%; Pred. No. 3.5e-204;
Matches 803; Conservative 0; Mismatches 55; Indels 4; Gaps 4;

Qy 503 TTAATACCAACCTCACTCGGCTCCATGAGGTCTC-AGCCCAAAACCCCTACTAACTGTTGG 561
Db 877 TAAACCAACCCCTCACTCGGCTCCATGAGTNCATCGGGCCAAAACCCCTACTAACTGTTGG 818
Qy 562 ATGTGCTCCCTCGCTGACCTTCAGGCCCATACATTTCAATCCCTGTTCTCTGAAACAATGGAAC 621
Db 817 ATATGCTCTCNCCTGAACTTCAGGGCATATGTTTCAATCCCTGTAAGCAATGGAAC 758
Qy 622 AACTTCAGACACAGA-AATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTGTTTCCAATCT 680
Db 757 ACTTTTCAGACACAGANAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTGTTTCCAATCT 698
Qy 681 GGAATAAACCCCATACCTCAAACTCAGCTGTGTAAATTTAGCAATATCTATAGACACAAC 740
Db 697 GGAATAAACCCCATACCTCAAACTCAGCTGTGT-AAAATTTAGCAATATCTATAGACACAAC 639
Qy 741 CAGCTCCCAATGATCAGGTGGGTAAACCTCCACACAGAAATAGTCTGCCCTACCTCAGG 800
Db 638 CAACTCNAATGATCAGGTGGGTAACTCTCCACACAANAATAGTCTGCCCTACCTCAGG 579
Qy 801 AATATTTTTGCTGTGTGTAACCTCAGCTATCATTTGTTGAATGGCTCTTTCAGAACTTAT 860
Db 578 AATATTTTTGCTGTGTGTAACCTCAGCTATCGTTGTTGAATGGCTCTTTCAGAACTTAT 519
Qy 861 GTGCTTCTCTCATTTCTTAGTGGCCCCCTATGACCATCTACCTGAAACAAGATTTATACAA 920
Db 518 GTGCTTCTCTCATTTCTTAGTGGCCCCCTATGACCATCTACCTGAAACAAGATTTATACAG 459
Qy 921 TCATGTCGTGTAACCTTAGCCCCCACAACAAAAGAGTACCCATCTTCTCTTTGTTATCAGAGC 980
Db 458 TTATGTCATATCTAAGCCCCCGCAAAAAGAGTAGTACCCATCTTCTCTTTGTTATAGAGC 399
Qy 981 AGGAGTGTAGGCAGACTAGGTACTGGCATTGGCAGTATCACAACTCTACTCAGTTCTA 1040
Db 398 AGGAGTGTAGGTGACACTAGGTACTGGCATTGGCGGTATCACAACTCTACTCAGTTCTA 339
Qy 1041 CTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCTCGTGTAC 1100
Db 338 CTACAAACTATCTCAAGAACTAAATGGGACATGGAACGGGTGCGGACTCCTCGTGTAC 279
Qy 1101 CTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCTTCAAAATCGAGAGCTTTAGA 1160
Db 278 CTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCTTCAAAATCGAGAGCTTTAGA 219
Qy 1161 CTTGCTAAACCGCAAAAGAGGGGAACCTGTTTATTTTATAGGAGAAGACGCTGTTATTA 1220
Db 218 CTTGCTAAACCGCTGAAAGAGGGGAACCTGTTTATTTTATAGGAGAAGATGCTGTTATTA 159
Qy 1221 TGTTAATCAATCCAGAAATGTCATGAGAAAGTTAAAGAAATTCGAGATCGAATACAAATG 1280
Db 158 TGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAAACG 99
Qy 1281 TAGACAGAGGAGCTTCAAAACACCCGAGCTGGGCGCTCTCAGCCCAATGGATGCCCTG 1340
Db 98 TAGACAGAGGAGCTTCAAAACACCCGAGCTGGGCGCTCTCAGCCCAATGGATGCCCTG 1340
Qy 1341 GGTTCCTCCCTCTTTAGGACCT 1362
Db 39 GATTCTCCCTCTTTAGGACCT 18

RESULT 13
BX388766/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BX388766 836 bp mRNA linear EST 29-APR-2004
BX388766 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1044YK06 3-PRIME, mRNA sequence.
BX388766
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 836)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 On May 8, 2003 this sequence version replaced gi:30462383.

Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 4215.r

For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?s=CS0AU012ZB04_U01082_l&c=4215.r.

FEATURES

Location/Qualifiers
 1. .836
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DI044YK06"
 /tissue_type="PLACENTA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 48.28; Score 714; DB 5; Length 836;
 Best Local Similarity 92.68; Pred. No. 6.5e-200;
 Matches 772; Conservative 0; Mismatches 60; Indels 2; Gaps 2;

565 TGCTCCCTCGCACTTCAGGCCATACATTTCAATCCCTGCTCTGCAACAATGGAACAAC 624
 835 TGCTCCCTCGCACTTCAGGCCCAATATTTCCCAATCCCTGCTCTGCAACAATGGAACAAC 776
 625 TTGAGCAGAGAAA-TAAACAGCACTTCGGTTTGTAGTAGGACCTTGTTCGAATCTGGA 693
 775 TTGAGCAGAGAAA-TACACACCACTCCCGTTTAAAGTAGGACCTTGTTCGAATCTGGA 716
 684 AATAACCCATACCTCAAACTCCTCTGTAAATTTAGCAATCTATAGACCAACCAAG 743
 715 AATAACCCATTCCTCAAACTCCTCTGTAAATTTAGCAATCTATAGACCAACCAAG 656
 744 CTCCCAATGCATCAGGTGGGTAACACCTCCACACGAATAGTCTGCCTACCCCTCAGGAAT 803
 655 CTCCCAATGCATCAGGTGGGTAACCTCTCCACACGAATAGTCTGCCTACCCCTCAGGAAT 596
 804 ATTTTGTCTGTGGTACCTCAGGCTATCATTTGTTGAATGGCTCTTCAGAACTATGTG 863
 595 ATTTTGTCTGTGGTACCTCAGGCTATCATTTGTTGAATGGCTCTTCAGAACTATGTG 536
 864 CTTCCTCTCATCTTAGTGCCCTATGACCATCTACACTGAACAAGATTTATACATCA 923
 535 CTTCCTCTCATCTTAGTGCCCTATGACCACTACACTGAACAAGATTTATACATCA 476
 924 TGTGCTACCTAAGCCCAACAAAGAGTAGTACCAATCTTTCCTTTTGTATCAGAGCAGG 983
 475 TGTATATCTAAGCCCAACAAAGAGTAGTACCAATCTTTCCTTTTGTATCAGAGCAGG 416
 984 AGTGTAGGAGAGACTAGTACTGGCATTTGGCAGTATCACAACTCTACTCAGTTCTACTA 1043
 415 AGTGTAGGAGAGACTAGTACTGGCATTTGGCAGTATCACAACTCTACTCAGTTCTACTA 356
 1044 CAAACTATCTCAAGAAATAAATGTGTGATCGGAACAGGTCACTGACTCCCTGGTCACTTT 1103
 355 CAAACTATCTCAAGAAATAAATGTGTGATCGGAACAGGTGCGCCACTCCCTGGTCACTTT 296
 1104 GCAAGATCAACTTAATCCCTTAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTTAGACTT 1163

Db 295 GCAAGATCAACTTAATCCCTTAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTTAGACTT 236
 Qy 1164 GCTAACCGCCAAAGAGAGGGGAAACCTGTTTATTTTATTTAGGAGAGAAACGCTGTTTATGTT 1223
 Db 235 GCTAACCGCTGAAGAGAGGGGAAACCTGTTTATTTTATTTAGGAGAGAAATGCTGTTATGTT 176
 Qy 1224 TAATCAATCCAGAAATGTCACGTAGAGAAATTTAAAGAAATTCAGATCGAATACAATGTAG 1283
 Db 175 TAATCAATCCGAAATCGTCACGTAGAGAAATTTAAAGAAATTCAGATCGAATACAATGTAG 116
 Qy 1284 ACAGAGAGCTTCAAAACACGACGCTGGGGCTCTCAGCCAATGATGCCCTGGGT 1343
 Db 115 ACAGAGAGGCTTCAAAACACGACGCTGGGGCTCTCAGCCAATGATGCCCTGGGT 56
 Qy 1344 TCTCCCTCTTTAGGACCTCTAGCAGCTCTAATATTTGTTTACTCTCTCTTTGGACC 1397
 Db 55 TCTCCCTCTTTAGGACCTCTAGCAGCTCTAATATTTGTTTACTCTCTCTTTGGGCC 3

RESULT 14

BX408734 921 bp mRNA linear EST 30-APR-2004
 LOCUS BX408734 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE012YJ24
 DEFINITION 5-PRIME, mRNA sequence.

ACCESSION BX408734
 VERSION BX408734.1 GI:30635959
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 921)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
 was not normalized. Library was constructed by Life Technologies, a
 division of Invitrogen.
 This sequence belongs to sequence cluster 4215.r
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?s=CS0BAA008ZE02_CS00686_2&c=4215.r

FEATURES

Location/Qualifiers
 1. .921
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DE012YJ24"
 /tissue_type="PLACENTA"
 /clone_lib="Homo sapiens PLACENTA"
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
 with a NotI-oligo(dT) primer. Five prime end enriched,
 double-strand cDNA was digested with Not I and cloned into
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."

ORIGIN

Query Match 47.9%; Score 709.8; DB 5; Length 921;
 Best Local Similarity 88.6%; Pred. No. 1.2e-198;
 Matches 815; Conservative 0; Mismatches 98; Indels 7; Gaps 4;

Qy 29 TTAAGTCTCTTCTTACCCCTTTCGCTCTCTCACTGACCCCTCCATGCTGCTGACACCA 88
 Db 1 TTTCTGTTCTTTTACCCCTTTCGCTCTCTCACTGACCCCTCCATGCTGCTGACCA 60
 Qy 89 GTAGCTCCCTTACCAAGAGTTTCTATGAAGAGCGGGCTTCTCTGGAATATTGATGCC 148

```
Db 61 GTAGTCCCTTACCAAGTTTCTATGGAGATGAGCGTCCGG-ATAATTGATGCC 119
Qy 149 CATCATATAGAGTTTATCTAAGGAAACTCCACCTTCACTGCCACACCCATATGCCCC 208
Db 120 CATCGTATAGAGTCTTCTAAGGAAACCCACCTTCACTGCCACACCCATATGCCCC 179
Qy 209 GCAACTGCTATACCTGCGCACTCTTTGCGATGATGCAATATCTATTGACAGGGA 268
Db 180 GCAACTGCTATACCTGCGCACTCTTTGCGATGATGCAATATCTATTGACAGGGA 239
Qy 269 AAATGATTAATCTAGTTGCTCGGAGGACTTGGAGCCACTGCTGTGGACTTACTTCA 328
Db 240 AAATGATTAATCTAGTTGCTCGGAGGACTTGGAGCTCTGTGTGGACTTACTTCA 299
Qy 329 CCATACCAAGTATGTCTGATGGGGTGGAAATCAAGGTGAGGCAAGAGAAAAACAAGTAA 388
Db 300 CCCAAACTGCTATGCTGATGGGGTGGAGTTCAAGATCAGGCAAGAGAAAAACATGTAA 359
Qy 389 AGGAAGCAATCTCCCACTGACCCGGGACATAGCACCCCTAGCCCTACAAAGGACTAG 448
Db 360 AAGAAGTAATCTCCCACTGACCCGGGTACATGGCACCTCTAGCCCTACAAAGGACTAG 419
Qy 449 TTCTCTCAAAACTACATGAACCTCGTACCACATCTCGCTGCTGGAGCTTATTTAATA 508
Db 420 ATCTCTCAAAACTACATGAACCTCGTACCACATCTCGCTGCTGGAGCTTATTTAATA 479
Qy 509 CCACTCTCACTCGGCTCCATGAGTCTCAGCCCCAAAACCTACTAACTGTGTGGATGCC 568
Db 480 CCACCCTCACTCGGCTCCATGAGTCTCGGCCCAAAAACCTACTAACTGTGTGGATGCC 539
Qy 569 TCCCTCTGCACTTCCAGCCCATACATTTCAATCCCTGCTGCTGAAACAATGGAACAATTCA 628
Db 540 TCCCTCTGCACTTCCAGCCCATATGTTTCAATCCCTGCTGCTGAAACAATGGAACAATTCA 599
Qy 629 GCACAGAAATAACACACCTTCCGTTTCTAGTAGGACTCTGTTTCCAACTCGGAATAA 688
Db 600 GCACAGAAATAACACACCTTCCGTTTCTAGTAGGACTCTGTTTCCAACTCGGAATAA 659
Qy 689 CCATACCTCAAACTCACCTGCTGTGTAAATTTAGCAATACTATAGACACACACGCTCCC 748
Db 660 CCATACCTCAAACTCACCTGCTGTGTAAATTTAGCAATACTATAGACACACCA-CTTCC 718
Qy 749 AATGATCAGGTGGGTAAACCTCCACAGCAATAGTCTGCCCTACCCCTAGGAATATTTT 808
Db 719 AATGATCAGGTGGGTAACTCTTCCCAAAATAGTCTGCCCTACCCCTCAAGGATATTT 778
Qy 809 TTGCTGTGTACCTCAGCCTATCATTTGTTTGAATGCTCTTCCAGATCTATGTGCTTCC 868
Db 779 TTGGCTGTGGGACCTCAGCCTATCGCGGTGATGGGCTTCCAGAA-ATAAGGGCTTCC 837
Qy 869 TCTCATTTCTTGTAGTGGCCCTTATGACCATCTACACTGAAACAAGATTTATACAATCATGTGC 928
Db 838 CCTCATCTTGTAGGCCCCCATGACCCCTTCACTGGAACAGA----TTATACAATTGTTCT 893
Qy 929 TACCTAAGCCCCACACAAA 948
Db 894 ATACTAAACCCCAACAAA 913
```

```
RESULT 15
BX347111/c
LOCUS
DEFINITION
  BX347111 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
  clone CS0D1022YE21 3-PRIME, mRNA sequence.
ACCESSION
  BX347111
VERSION
  BX347111.1 GI:30375145
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 844)
```

AUTHORS
TITLE
JOURNAL
COMMENT

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
7108.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?S=CS0AU0102B08_U0922_1&c=7108.r.

FEATURES
source

1..844
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1022YE21"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 47.7%; Score 705.8; DB 5; Length 844;
Best Local Similarity 92.7%; Pred. No. 1.8e-197;
Matches 762; Conservative 0; Mismatches 58; Indels 2; Gaps 2;
Qy 502 TTTAATACCAACCTCACTCGGCTCCATGAGTCTCAGCCCAAAA-CCCTACTAACTGTTG 560
Db 837 TTAATCCACCTCCCTCGGTTCCATTAGGTTCTCGGCCCAAAACCCCTACTAACTGTTG 778
Qy 561 GATGTGCTT-CCGCTGCACTTCAGGCCATACATTTCAATCCCTGTTCTGGAACAATGGA 619
Db 777 GATATGCTTCCGCTGAAATTTTTCAGGCCATATGTTTCAATCCCTGTTACCTGAACAATGGA 718
Qy 620 ACAACTTTTCAGCAGAGAAATAAACACCACTTCGTTTGTAGTAGGACCTCTGTTTCCAATC 679
Db 717 CAACTTTTCAGCAGAGAAATAAACACCACTTCGTTTGTAGTAGGACCTCTGTTTCCAATC 658
Qy 680 TGAATAAATACCCATACCTCAAACTCACCTGTGTAAATTTAGCAATACTATAGACACAA 739
Db 657 TGAATAAATACCCATACCTCAAACTCACCTGTGTAAATTTAGCAATACTATAGACACAA 598
Qy 740 CCAAGTCCCAATGCAATCAGGTGGGTAAACCTCCACAGCAATAGTCTGCTTACCTCAG 799
Db 597 CCAACTCNAATGCAATCAGGTGGGTAACTCTCCACACAAATAGTCTGCTTACCTCAG 538
Qy 800 GAATATTTTTCGTCGTGCTACTCTCAGCCTATCATTTGTTGATGCTCTTTCAGAACTTA 859
Db 537 GAATATTTTTCGTCGTGCTACTCTCAGCCTATCATTTGTTGATGCTCTTTCAGAACTTA 478
Qy 860 TGTGCTTCTCTCATTTCTTGTAGTGGCCCTTATGACCATCTACACTGAAACAAGATTTATACA 919
Db 477 TGTGCTTCTCTCATTTCTTGTAGTGGCCCTTATGACCATCTACACTGAAACAAGATTTATACA 418
Qy 920 ATCATGCTGTAATAGCCCCCAACAAAGAGTACCCATTTCTTCTTTGTTGTTATCAGAG 979
Db 417 GTTATGTCATATCTAAGCCCGCAACAAAGAGTACCCATTTCTTCTTTTGTATAGGAG 358
Qy 980 CAGGAGTGTGACGACTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1039
Db 357 CAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 298
Qy 1040 ACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCAGTCTCCCTGGTCA 1099
Db 297 ACTACAAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCAGTCTCCCTGGTCA 238

QY	1100	CTTGTCAAGATCAACTAACTCCCTAGCAGCAGTACTCTCTCAAAATCGAAGAGCTTTAG	1159
Db	237	CTTGTCAAGATCAACTAACTCCCTAGCAGCAGTACTCTCTCAAAATCGAAGAGCTTTAG	178
QY	1160	ACTTGTCAACCGCAAAAGAGGGGGAACCTGTTTATTTTAGGAGAAAGACCTGTTATT	1219
Db	177	ACTTGTCAACCGCTGAAGAGGGGGAACCTGTTTATTTTAGGAGAAAGATGCTGTTATT	118
QY	1220	ATGTTTAATCAATCCAGAAATGTCACATGAGAAAGTTAAAGAAATTCGAGATCGAATACAAT	1279
Db	117	ATGTTTAATCAATCCGAATCGTCACATGAGAAAGTTAAAGAAATTCGAGATCGAATACAAC	58
QY	1280	GTAGAGCAGAGGAGCTTCAAAACACCGAACGCTGGGGCCCTCC	1321
Db	57	GTAGAGCAGAGGAGCTTGGAAACACTGGACCTGGGGCCGCC	16
RESULT 16			
LOCUS	BX367907	828 bp	linear EST 28-APR-2004
DEFINITION	BX367907 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA		
ACCESSION	BX367907		
VERSION	BX367907.1	GI:30459586	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1	(bases 1 to 828)	
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.		
TITLE	Full-length cDNA libraries and normalization		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 4215.r		
FEATURES	source		
	1..828		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CSODI051YM13"		
	/tissue_type="PLACENTA COT 25-NORMALIZED"		
	/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"		
	/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."		
ORIGIN			
Query Match	47.5%;	Score 703.2;	DB 5; Length 828;
Best Local Similarity	93.8%;	Pred. No. 1e-196;	
Matches	754;	Conservative 0;	Mismatches 48; Indels 2; Gaps 2;
QY	110	TTCTATGAGAACGGCGTCTCTGGAATATTGATGCCCATCATATAGGAGTTTATCTA	169
Db	9	TTTCTTGAGAAATGAGCGTCCCGG-AATATTGATGCCCATCATATAGGAGTCTTCTA	67
QY	170	AGGGAATCCACTTCACTGCCACACCCATATGCCCCGCACTGCTATTAACCTCGCA	229
Db	68	AGGGAACCCCACTTCACTGCCACACCCATATGCCCCGCACTGCTATTAACCTCGCA	127
QY	230	CTCTTTCATCGCAATCACTATTATTCGACAGGGAATGATTAATCTTAGTTGTC	289

Db	128	CTCTTTGCATGCATCAAACTACTATTATTGGACAGGAAAAATGATTAACTCTAGTTGTC	187
QY	290	CTGGAGGACTTGGAGCCCACTGTCTGTGGACTTACTTCAACCATACAGATATGTCTGATG	349
Db	188	CTGGAGGACTTGGAGCTCACTGTCTGTGGACTTACTTCAACCATACAGATATGTCTGATG	247
QY	350	GGGGTGGAAATCAAGGTGAGGAGAGAGAAACAGTAAGGAAGCAATCTCCCAACTGA	409
Db	248	GGGGTGGAGTTCAAGATCAGGCAAGAGAAAAACATGTAAAGAAAGTAATCTCCCAACTCA	307
QY	410	CCCGGGGACATAGCACCCTAGCCCTCAAAAGGACTAGTTCTCTCAAAATACATGAAA	469
Db	308	CCCGGGTACATGACCTCTAGCCCTCAAAAGGACTAGTTCTCTCAAAATACATGAAA	367
QY	470	CCCTCCGTACCCATCTCGCTGGTGGAGCCTTAATTAATACACCCCTCACTGGCTCCATG	529
Db	368	CCCTCCGTACCCATCTCGCTGGTGGAGCCTTAATTAATACACCCCTCACTGGCTCCATG	427
QY	530	AGGTCTCAGCCCAAAACCTACTACTTGGATGTCCTCCCTCCCTGCACTTCAGGCCAT	589
Db	428	AGGTCTCAGCCCAAAACCTACTACTTGGATGTCCTCCCTCCCTGCACTTCAGGCCAT	487
QY	590	ACATTTCAATCCCTGTTCTGAAACAATGGAACAACCTTCAGCAAGAAAATAAACACACTT	649
Db	488	ATGTTTCAATCCCTGTTCTGAAACAATGGAACAACCTTCAGCAAGAAAATAAACACACTT	547
QY	650	CGTTTTAGTAGACCTCTTGTTCATCTCGAATAATCCCATACCTCAAAACCTCACT	709
Db	548	CGTTTTAGTAGACCTCTTGTTCATCTCGAATAATCCCATACCTCAAAACCTCACT	607
QY	710	GTGTAATAATTTAGCAATATATAGACACAAACAGCTCCCAATGCATCAGTGGGTAAACAC	769
Db	608	GTGTAATAATTTAGCAATATATAGACACAAACAGCTCCCAATGCATCAGTGGGTAACTC	667
QY	770	CTCCACACAGAAATAGTCTGCTCCTCAGCAATATTTTCTGTGTGTACTCAGCCT	829
Db	668	CTCCACACAAATAGTCTGCTCCTCAGCAATATTTTCTGTGTGTACTCAGCCT	727
QY	830	ATCATTGTTTGAATGGCTCTCAGATCTATGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	889
Db	728	ATCGTTGTTTGAATGGCTC-TCAGAAATATATGGCTCTCTCTCTCTCTCTCTCTCTCTCT	786
QY	890	TGACCATCTACACTGAAACAAGATT	913
Db	787	TGACCATCTACACTGAAACAAGATT	810
RESULT 17			
LOCUS	BX368078/c	846 bp	mRNA linear EST 26-APR-2004
DEFINITION	BX368078 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA		
ACCESSION	BX368078		
VERSION	BX368078.2	GI:46572908	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1	(bases 1 to 846)	
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.		
TITLE	Full-length cDNA libraries and normalization		
JOURNAL	Unpublished (2001)		
COMMENT	On May 8, 2003 this sequence version replaced gi:30447710. Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster		

4215.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0A0U012ZD05_U01092_1&c=4215.r.

FEATURES
source
1. .846
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1051YML3"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 47.2%; Score 698.6; DB 5; Length 846;
Best Local Similarity 92.8%; Pred. No. 2.4e-195;
Matches 786; Conservative 0; Mismatches 56; Indels 5; Gaps 5;

QY 467 AAACCTCCGTACCCACTACGCTGGTGAGCCTATTATATACCAC-CCTCACTCGGCTC 525
DB 846 AAACCTCCGGTACCCACTACGCTGGTGTAGCCTATTATATACAAACCTCACTGGGCTC 787
QY 526 CATGAGGTCTCAGCCC-AAACCTTACTTAATCTGTGGATGGCTGCCCTGCACCTCA- 583
DB 786 ATATAGGTCTCGGCCCAAAACCTTACTTAATCTGGAAATATGCTCCCTCGAACTTCAG 727
QY 584 GGCCATACATTTCAT-CCCTGTTCTCGAACAATGGAACAATTCAGCAGAGAATAAACC 642
DB 726 GGCCATATGTTTCAATNCCCTGACCTGACATGGAACAATTCAGCAGAGAATAAACC 667
QY 643 ACCACTTCCGTTTTAGTAGGACCTCTGTTTCCAAATCTGGAATAACCCATACCTCAAAC 702
DB 666 ACCACTNCCGTTTTAGTAGGACCTCTGTTTCCAAATCTGGAATAACCCATACCTCAAAC 607
QY 703 CTCACTGTGTAAATTTAGCAATATCTATAGACACACACAGCTCCCAATGCATCAGGTGG 762
DB 606 CTCACTGTGTAAATTTAGCAATATCTATAGACACACACCACTCCCAATGCATCAGGTGG 547
QY 763 GTAACACCTCCACACGAATGCTGCCTACCTCCAGGAATATTTTGTCTGTGTAC 822
DB 546 GTAACCTCCACACGAATGCTGCCTACCTCCAGGAATATTTTGTCTGTGTAC 487
QY 823 TCAGCCTATCATTTGTTGATGCTCTTCAGAACTATATGCTTCCTCTCATTTCTTAGTG 882
DB 486 TCAGCCTATGTTGTTGATGCTCTTCAGAACTATATGCTTCCTCTCATTTCTTAGTG 427
QY 883 CCCCCTATGACCATCTACACGAAAGATTATACAAATCATGTCGTACCTAAGCCCCAC 942
DB 426 CCCCCTATGACCATCTACACGAAAGATTATACAAATCATGTCGTACCTAAGCCCCAC 367
QY 943 AACAAAGAGTACCCATCTCTCTCTTTGTTATCAGAGCAGGAGTGTGAGCAGCTAGGT 1002
DB 366 AACAAAGAGTACCCATCTCTCTCTTTGTTATAGGAGCAGGAGTGTGAGTGCACCTAGGT 307
QY 1003 ACTGGCAATGGCAGTATCAAACTCTACTCAGTTCTACTACAACTCTCTCAAGAAATA 1062
DB 306 ACTGGCAATGGCAGTATCAAACTCTACTCAGTTCTACTACAACTCTCTCAAGAAATA 247
QY 1063 AATGGTGACATGGAACAGGTCTACTGACTCCCTGGTCACTTGCAGAGATCAACTTAACTCC 1122
DB 246 AATGGTGACATGGAACAGGTCTACTGACTCCCTGGTCACTTGCAGAGATCAACTTAACTCC 187
QY 1123 CTAGCAGCAGTACTCTTCAAAATCGAAGCTTTAGACTTGTCTAACCCGCAAGAGGG 1182
DB 186 CTAGCAGCAGTACTCTTCAAAATCGAAGCTTTAGACTTGTCTAACCCGCTGAAGAGGG 127
QY 1183 GGAACCTGTTTATTTTATAGGAGAGAACGCTGTTTATTTATGTTTAACTCAATCGAAATGTC 1242
DB 126 GGAACCTGTTTATTTTATAGGAGAGAACGCTGTTTATTTATGTTTAACTCAATCGGATCGTC 67

QY 1243 ACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAAATGTAGACAGAGAGCTTCAAAC 1302
DB 66 ACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAAAGT-GAGCAGAGAGAGCTTCAAAC 8
QY 1303 ACCGAAC 1309
DB 7 ATGACC 1
RESULT 18
BX389657
LOCUS
DEFINITION
BX389657 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1051YML3 5-PRIME, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 995)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 8, 2003 this sequence version replaced gi:30462931.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
4215.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0BAF0272B05_AF02531_2&c=4215.r

FEATURES
source
1. .995
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1051YML3"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 47.0%; Score 696.2; DB 5; Length 995;
Best Local Similarity 86.8%; Pred. No. 1.3e-194;
Matches 826; Conservative 0; Mismatches 114; Indels 12; Gaps 5;

QY 266 GGAATATGATTAATCCTAGTTGCTCGAGGAGCTTGAGGACCTGCTGTGTGACTTACT 325
DB 1 GGAATATGATTAATCCTAGTTGCTCGAGGAGCTTGAGGAGT-ACTGCTGTGTGACTTACT 59
QY 326 TCACCCATACCAAGTATGTCGTATGTTGGGGTGGAAATCAAGGTCAAGCAAGAAAAACAAG 385
DB 60 TCACCCAACTCGTATGTCGTATGTTGGGGTGGAGTTCAAGATCAGGCAAGAAAAACAATG 119
QY 386 TAAAGGAACAATCTCCCACTGACCCGGGACATAGCAGCCCTAGCCCTACAAAGGAC 445
DB 120 TAAAGGAAGTAATCTCCCACTGACCCGGGTACATGGCACTCTAGCCCTACAAAGGAC 179
QY 446 TAGTCTCTCAAAACTACATGAACCCCTCGGTACCCATACCTCGCTCGCTGGTGGAGCTATTTA 505
DB 180 TAGATCTCTCAAAACTACATGAACCCCTCGGTACCCATACCTCGCTCGCTGGTGGAGCTATTTA 239

```
QY 506 ATACCACTCTCACTCGCTCCATGAGTCTCAGCCCAAAACCTACTAACTGTGTGGATGT 565
DB 240 ATACCACTCTCACTCGCTCCATGAGTCTCAGCCCAAAACCTACTAACTGTGTGGATAT 299
QY 566 GCTCTCCCTGCACTTCAGGCCATACATTTCAATCCCTGTTCTTGAACATGGAACAAC 625
DB 300 GCTCTCCCTGCACTTCAGGCCATATGTTTCAATCCCTGTTCTTGAACATGGAACAAC 359
QY 626 TCAGCAGAGAAATAAACAACACCTTCGTTTCTAGTAGGACCTCTGTTTCCAACTCTGAAA 685
DB 360 TCAGCAGAGAAATAAACAACACCTTCGTTTCTAGTAGGACCTCTGTTTCCAACTCTGAAA 419
QY 686 TAACCCATACCTCAAACTCACCTGTGTAAATTTAGCAATACTATAGACACAAACAGCT 745
DB 420 TAACCCATACCTCAAACTCACCTGTGTAAATTTAGCAATACTATAGACACAAACAGCT 479
QY 746 CCAATGTCATCAGGTGGTAAACCTCCACAGATAGTCTGCTTACCTCAGGATAT 805
DB 480 CCAATGTCATCAGGTGGTAAACCTCCACAGATAGTCTGCTTACCTCAGGATAT 539
QY 806 TTTTCTCTGTGTGCTCAGCTATCATTTGTTGATGCTCTTCTCAGAACTCTATGTGCT 865
DB 540 TTTTCTCTGTGTGCTCAGCTATCATTTGTTGATGCTCTTCTCAGAACTCTATGTGCT 599
QY 866 TCCTCTCATTTCTAGTGTCCCTTACCATCTACATGACAGAAAGATTTATACAACTCATG 925
DB 600 TGCTCTATTTCTAGTGTCCCTTACCATCTACATGACAGAAAGATTTATACAACTCATG 659
QY 926 TCGTACCTAAGCCGCCAACAAGAGTACCATCTTCTTCTTCTTGTATCAGAGCAGGAG 985
DB 660 TCATATCTAAGCCGCCAACAAGAGTACCATCTTCTTCTTGTATCAGAGCAGGAG 719
QY 986 TGCTAGCAGACTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1045
DB 720 TGCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 779
QY 1046 AACTATCTCAA-GAAATAAATGTGACATGGAACAGGT---CACTGATCCCTGGTCACC 1101
DB 780 AACTATTTAAGGACCTAAATGGAACATGGAACGGGTGCGGACCTCCCTGGTAACT 839
QY 1102 TTGCAAGATCACTTAATCTCC---TAGCAGCGTAGTCTTCAAAATCGAAGAGCTTT 1157
DB 840 TGAAGATCACTTAATCTCC---TAGCAGCGTAGTCTTCAAAATCGAAGAGCTTT 899
QY 1158 AGACTCTCAACGCG---CAAAAGAGGGGAACCTGTTATTTTAGGAGAA 1206
DB 900 AGACTCTCAACGCGCTGTAAGAGGGGGGAACCTGTTATTTTAGGAGAA 951

RESULT 19
BX347314
LOCUS
DEFINITION BX347314 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1044YK06 5-PRIME, mRNA sequence.
ACCESSION BX347314
VERSION BX347314.1 GI:30365192
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 924)
JOURNAL Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
COMMENT Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life technologies, a
```

```
division of Invitrogen. This sequence belongs to sequence cluster
4215.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0BAA016ZH06_CS01489_1&c=4215.r

FEATURES
source
Location/Qualifiers
1..924
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1044YK06"
/cisue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 46.9%; Score 695; DB 5; Length 924;
Best Local Similarity 92.7%; Pred. No. 2.9e-194;
Matches 750; Conservative 0; Mismatches 57; Indels 2; Gaps 2;
QY 1 ATGGCCCTCCCTTATCATACATCTTTCTTTTACTGTCTCTTACCCCTTTCCCTCTCACT 60
DB 112 ATGGCCCTCCCTTATCATATTTTCTTTTACTGTCTTTTACCTCTTTTACCTCTCACT 171
QY 61 GCACCCCTCCCTTATCATACATCTTTCTTTTACTGTCTCTTACCCCTTTCCCTCTCACT 120
DB 172 GCACCCCTCCCTTATCATACATCTTTCTTTTACTGTCTTTTACCTCTTTTACCTCTCACT 231
QY 121 AGCGGCTCTCTGGAATATTTGATGCCCATATATAGGAGTTTATCTAAGGGAACCTCC 180
DB 232 ATGACGCTCCCGGAATATTTGATGCCCATATATAGGAGTTTATCTAAGGGAACCTCC 291
QY 181 ACCTTCACTGCCACACCCATATGCCCCGCACTGCTATAAATCTCCCACTCTTTCGATG 240
DB 292 ACCTTCACTGCCACACCCATATGCCCCGCACTGCTATAAATCTCCCACTCTTTCGATG 351
QY 241 CATGCAATTAATCTATTTTGGACAGGGAATATGATTAATCTTCTAGTTGTCTTGGAGACTT 300
DB 352 CATGCAATTAATCTATTTTGGACAGGGAATATGATTAATCTTCTAGTTGTCTTGGAGACTT 411
QY 301 GGAGCCTCTCTGTTGGACTTACTTCACTACCATACCAGTATGTCATGGGGTGGAAAT 360
DB 412 GGAGTCACTCTGTTGGACTTACTTCACTACCATACCAGTATGTCATGGGGTGGAGTT 471
QY 361 CAGGTCAGCAGAGAGAAACAAAGTAAAGGAAGCAATCTCCCACTGACCCGGGACAT 420
DB 472 CAAGATCAGGCAAGAGAAACAAAGTAAAGGAAGTAAATCTCCCACTCACCCTGGGTACAT 531
QY 421 AGCACCCCTAGCCCTTACAAAGGACTAGTTCTCTCAAACTACATGAAACCTCCCGTACC 480
DB 532 GGCACTCTAGCCCTTACAAAGGACTAGATCTCTCAAACTACATGAAACCTCCCGTACC 591
QY 481 CATACTCGCTGTGTGAGCTTATTTATACACCTCTACTCGGCTCCATGAGTCTCAGCC 540
DB 592 CATACTCGCTGTGTGAGCTTATTTATACACCTCTACTCGGCTCCATGAGTCTCAGCC 651
QY 541 CAAACCCCTACTAACTGTGTGATGTGCTCCCTGCACTTTCAGGCGCATACATTTCAATC 600
DB 652 CAAACCCCTACTAACTGTGTGATGTGCTCCCTGCACTTTCAGGCGCATACATTTCAATC 710
QY 601 CTTGTTCTCGAACAATGGAACAACTTTGAGCAGACAGAAATAAACCCTTCCGTTTATGTA 660
DB 711 CTTGTTCTCGAACAATGGAACAACTTTGAGCAGACAGAAATAAACCCTTCCGTTTATGTA 770
QY 661 GGACCTCTCTTCTTCCATCTGGAATACCATACCTCAAACTCAGCTCTGTAATTTT 720
DB 771 GGACCTCTCTTCTTCCATCTGGAATACCATACCTCAAACTCAGCTCTGTAATTTT 830
QY 721 AGCAATACTATAGACACAAACAGCTCCCAATGATCAGGTGGGTAAACCTCCCAACGA 780
```

```

Db      831 AGCATACTATACATACACCAAAATCCCATGATCAAGGTGGTAACTCTCCACAC-A 889
QY      781 ATAGTCTGCTACCTCAGGAATATTTT 809
Db      890 ATAGTCTGCTACCTCAGGAATATTTT 918

RESULT 20
LOCUS   AU138405
DEFINITION AU138405 PLACE1 Homo sapiens cDNA clone PLACE1008489 5', mRNA
ACCESSION AU138405
VERSION   AU138405.1 GI:10999926
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          1 (bases 1 to 723)
          Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,
          Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and
          Isogai, T.
          HRI human cDNA project
          Unpublished (2000)
          Contact: Takao Isogai
          Genomics Laboratory
          Helix Research Institute
          1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
          Tel: 81-438-52-3975
          Fax: 81-438-52-3986
          Email: genomics@hri.co.jp
          HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
          Research Institute; cDNA library construction: Department of
          Virology, Institute of Medical Science, University of Tokyo, and
          Helix Research Institute.
          Location/Qualifiers
            source          1. .723
                           /organism="Homo sapiens"
                           /mol_type="mRNA"
                           /db_xref="taxon:9606"
                           /clone="PLACE1008489"
                           /tissue_type="placenta"
                           /clone_lib="PLACE1"
                           /notes="Vector: pME18SFL3"

FEATURES
source          1. .723
                           /organism="Homo sapiens"
                           /mol_type="mRNA"
                           /db_xref="taxon:9606"
                           /clone="PLACE1008489"
                           /tissue_type="placenta"
                           /clone_lib="PLACE1"
                           /notes="Vector: pME18SFL3"

ORIGIN
Query Match      43.7%; Score 647; DB 1; Length 723;
Best Local Similarity 94.6%; Pred. No. 4.7e-180;
Matches 679; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY      565 TGCTCCCTGCACTTCAGGCCATACATTTCAATCCCTGTTCTGAAACAATGGAACAAC 624
Db      3   TGCTCCCTGCACTTCAGGCCATATGTTTCAATCCCTGTTCTGAAACAATGGAACAAC 62
QY      625 TTGAGCAGAGAAATAAACACCACTTCGGTTTATAGGACCTCTTGTTCGAATCTGGAA 684
Db      63   TTGAGCAGAGAAATAAACACCACTTCGGTTTATAGGACCTCTTGTTCGAATCTGGAA 122
QY      685 ATAAACCATCTCAACCTCAGCTGTGTAAATTTAGCAATCTATAGACACAACGCG 744
Db      123 ATAAACCATCTCAACCTCAGCTGTGTAAATTTAGCAATCTATAGACACAACGCG 182
QY      745 TCCCAATGCAATCAGGTGGGTAACACCTCCACACGAAATAGTCGCTACCTCAGGAATA 804
Db      183 TCCCAATGCAATCAGGTGGGTAACCTCCCAACAATAGTCGCTACCTCAGGAATA 242
QY      805 TTTTGTCTGTGTGCTACCTCAGCCTATCATTTGTTGAATGGCTCTTCAGAAATCTATGTGC 864
Db      243 TTTTGTCTGTGTGCTACCTCAGCCTATCATTTGTTGAATGGCTCTTCAGAAATCTATGTGC 302
QY      865 TTCTCTCATCTTGTAGTGGCCCTATGACCATCTACACTGAAACAAGTTTATACATCAT 924

```

```

Db      303 TTCTCTCATCTTCTAGTGGCCCTATGACCATCTACACTGAACAAGTTTATACAGTTAT 362
QY      925 GTGCTACTAAGCCCCACAAAGAGTACCCATCTTCTCTTTTGTATCAGAGCAGGA 984
Db      363 GTCATATCTAAGCCCGCAACAAAGAGTACCCATCTTCTCTTTTGTATAGGAGCAGGA 422
QY      985 GTGCTAGGCGAGCTAGGTAGTGGCATTGGCAGTATCACACCTCTACTCAGTTCTACTAC 1044
Db      423 GTGCTAGGTGCTACTAGGTAGTGGCATTGGCAGTATCACACCTCTACTCAGTTCTACTAC 482
QY      1045 AAATCTATCTCAAGAAATAAATGTTGATGCAATGCAACAGGTCACTGACCTCCCTGGTCACTTG 1104
Db      483 AAATCTATCTCAAGAAATAAATGTTGATGCAATGCAACAGGTCACTGACCTCCCTGGTCACTTG 542
QY      1105 CAAGATCAACTTAACTCCCTAGCAGCAGTAGTCTCTTCAAAATCGAAGAGCTTTAGACTTG 1164
Db      543 CAAGATCAACTTAACTCCCTAGCAGCAGTAGTCTCTTCAAAATCGAAGAGCTTTAGACTTG 602
QY      1165 CTAAACCGCAAGAGAGGGGAACTGTTATTTTATAGGAGAAACGCTGCTTATATGTT 1224
Db      603 CTAAACCGCTGAAGAGAGGGGAACTGTTATTTTATAGGAGAAACGCTGCTTATATGTT 662
QY      1225 AATCAATCCAGAAATTCCTACT-GAGAAAGTTAAAGAAATTCGAGATCGAATACAAATGT 1281
Db      663 AATCAATCCGGGATCGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAAATGT 720

RESULT 21
LOCUS   AQ261133
DEFINITION AQ261133
ACCESSION AQ261133
VERSION   AQ261133.1 GI:3787657
KEYWORDS GSS.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          1 (bases 1 to 631)
          Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
          Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and
          Venter, J.C.
          Use of a random human BAC End Sequence Database for Sequence-Ready
          Map Building
          Unpublished (1998)
          Other GSSs: CITBI-E1-2506C15.TR
          Contact: Mark Adams
          Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850, USA
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: mdadams@tigr.org
          Clones are available from Research Genetics (info@resgen.com). BAC
          end search page:
          http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
          Seq primer: M13-21
          Class: BAC ends.

FEATURES
source          1. .631
                           /organism="Homo sapiens"
                           /mol_type="genomic DNA"
                           /db_xref="taxon:9606"
                           /clone="2506C15"
                           /sex="male"
                           /cell_type="sperm"
                           /clone_lib="CITBI-E1"
                           /note="Vector: pBelBAC11; Site_1: EcoRI; Site_2: EcoRI;
                           CalTech Human BAC Library D"

ORIGIN
Query Match      42.5%; Score 629.4; DB 8; Length 631;

```



```

ACCESSION BX409328
VERSION BX409328.2 GI:46932867
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 924)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 13, 2003 this sequence version replaced gi:30652931.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 9106 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
This sequence belongs to sequence cluster 4215.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0BAE0122F02_AE01071_1&c=4215.r

FEATURES
source
Location/Qualifiers
1..924
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0BE013Y120"
/tissue_type="PLACENTA"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN
Query Match 42.1%; Score 623.8; DB 5; Length 924;
Best Local Similarity 85.5%; Pred. No. 3.8e-173;
Matches 748; Conservative 0; Mismatches 120; Indels 7; Gaps 5;

QY 1 ATGGCCCTCCCTTATCATATCTTTCTTTACTGTTCTCTTACCCCTTTGCGTCTCACT 60
Db 111 ATGGCCCTCCCTTANATATTTTCTTTACTGTTCTTTTACCCCTTTTCACTCTCACT 110
QY 61 GCACCCCTCCATGCTGCTGACACAGCAGTAGTCCCTTACCAAGAGTTTCTATGAAGA 120
Db 111 GCACCCCTCCATGCGCGCTGTATGACCAAGTAGTCCCTTACCAAGAGTTTCTATGGAGA 170
QY 121 AGCGGCTTCTCGGAATATTGATGCCCATATATAGGAGTTTATCTAAGGGAAGTCC 180
Db 171 ATGCAGGTGCCGGAATATTGATGCCCATATCTATAGGAGTCTTCTAAGGGAAGTCC 230
QY 181 ACCTTCACTGCCACACCCATATGCCCGCAACTGCTATTAACCTCTGCCACTCTTTGCATG 240
Db 231 ACCTTCACTGCCACACCCATATGCCCGCAACTGCTATCACTCTGCCACTCTTTGCATG 290
QY 241 CATGCAAACTACTATTTATGGACAGGGAATAATGATTAATCTAGTTGCTCTGGAGACTT 300
Db 291 CATGCAAACTACTATTTATGGACAGGGAATAATGATTAATCTAGTTGCTCTGGAGACTT 350
QY 301 GGAGCCACTGCTGTTGGACTTACTTCCACCATACCACTATGCTCATGGGGTGAATT 360
Db 351 GGAGTCACTGCTGTTGGACTTACTTCCACCATACCACTATGCTCATGGGGTGAATT 410
QY 361 CAAAGGTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 420
Db 411 CAAAGTACGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 470
QY 421 AGCACCCCTAGCCCTTACAAAGGACTAGTTTCTCTCAAAACTACATGAAACCCCTCCGTACC 480

```

```

Db 471 GGCACCTCTAGCCCCCTACAAAGACTAGATCTCTCAAAACTACATGAAACCCCTCCGTACC 530
QY 481 CATACTCGCTCGTGGAGCCTATTTAATACCAACCCCTCACTCGGCTCCATGAGGTCTCAGCC 540
Db 531 CATACTCGCTCGTGAAGCCTATTTAATACCAACCCCTCACTCGGCTCCATGAGGTCTCGGCC 590
QY 541 CAAAACCCCTACTAACTGTTGGATGTCCTCCCTCGCTTTCAGGCGCATACATTTCATC 600
Db 591 CANAACCCCTACNNTACTGTGGATATGCTCCCTCCCTGACTTANNGCATATNGTT--CATC 648
QY 601 CTTGTTCTCTGACAATGGAACAACTTTCAGACAGAAATAAACACCACTTCCGTTTGTAGTA 660
Db 649 CTTGTACCTG-ACATGGAACACTTTCAGACAGAAATAAACACCACTTTCGTTTGTAGTA 707
QY 661 GGACCTCTTTGTTTCCAACTCTGGAATAAACCCATACCTCAAACTCACTGTGTGTAATTT 720
Db 708 GGACCTCTTTGTTTCCATCTGGG--ATTACCCATACCTAAACCTNACCTGGGT-AAATTT 764
QY 721 AGCAATACTATAGACACAACACGCTCCCAATGATCAGGTGGGTAAACCTCCACACGA 780
Db 765 AGCATTACTACTTCCCAACCACTTCCCATGATCAGGTGGGTAGCTCTTCCCAACCAA 824
QY 781 ATAGTCTGCTACCCCTCAGGAATATTTTGTCTGTGTACCTCAGCCTATCATTTGTTG 840
Db 825 ATAGTCTGCTACCCCTCAGGA--ATTTTGTGTGTGGACCTTAACCTTTGTTGTTG 883
QY 841 AATGGCTCTTCAAGATCTATGTCTTCTCTCTCAT 875
Db 884 AAAGGCTCTTAAAGAACTATAGGCCCTCTCTCAAT 918

RESULT 24
BE734284
LOCUS 714 bp mRNA linear EST 15-SEP-2000
DEFINITION 601565487F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3840572 5',
mRNA sequence.
ACCESSION BE734284
VERSION BE734284.1 GI:10148276
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 714)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM530 row: m column: 21
High quality sequence stop: 712.
Location/Qualifiers
1..714
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3840572"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 21"
/notes="Organ: placenta; Vector: pOTB7; Site:1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University

```

[illegible]

RESULT	25
LOCUS	BX430050/c
DEFINITION	BX430050 Homo sapiens PLACENTA Homo sapiens CDNA clone CS0DE012YU24 658 bp mRNA linear EST 04-MAY-2004
ACCESSION	BX430050
VERSION	BX430050
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 658) Li,W.B., Gruber,C., Jessee,J. and Polayes,D. Full-length cDNA libraries and normalization Unpublished (2001)
AUTHORS	
TITLE	
JOURNAL	

On May 15, 2003 this sequence version replaced gi:30776872.

Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized. Library was constructed by Life Technologies, a division of Invitrogen.

This sequence belongs to sequence cluster 4215.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0BA5002D11_CS00181_1&c=4215.r

Location/Qualifiers

1. 658
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE012YJ24"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

FEATURES
source

ORIGIN

Query Match	39.2%;	Score 580.4;	DB 5;	Length 658;
Best Local Similarity	93.6%;	Pred. No. 2.6e-160;		
Matches 616;	Conservative 0;	Mismatches 41;	Indels 1;	Gaps 1;
Qy	505	AATACCA	CCCTCACTCGGCTCCATGAGGTC	CTCAGGCCAAA-ACCCTACTAACTGTTGGAT 563
Db	658	AAAACCA	TCCCAATGGATACCATAGGTTT	TGGTACAAATACCTACTAACTGTTGAAT 599
Qy	564	GTCCCTCC	CCCTGGCATCTTCAGGCGATACAT	TTTCAATCCCTGTTCCCTGAAACAATGGAAACAA 623
Db	598	ATGCTCC	CCCTGAACTTCAGGCGCATATGTT	CCAATCCCTGTACTCTGAAACAATGGAAACAA 539
Qy	624	CTTCAGC	ACAGAAATAAACACCACTTC	CCGTTTGTAGTAGGACCTCTTGTTCCCAATCTGGA 683
Db	538	CTTCAGC	ACAGAAATAAACACCACTTC	CCGTTTGTAGTAGGACCTCTTGTTCCCAATCTGGA 479
Qy	684	AATAACCC	TACCTCAAAACCTCACTGTGT	ATAAATTTAGCAATACTATAGACACAAACAG 743
Db	478	AATAACCC	TACCTCAAAACCTCACTGTGT	ATAAATTTAGCAATACTATACATACACAAACAA 419
Qy	744	CTCCAA	TGCAATCAGGTGGGTAAACCT	CCCAACAGAAATAGTCCTGCTACCCCTCAGGAAT 803
Db	418	CTCCAA	TGCAATCAGGTGGGTAACT	CTCCCAACAAATAGTCCTGCTACCCCTCAGGAAT 359
Qy	804	ATTTTT	TGTCTGTGGTACCTCAGGCT	ATCATGTTTGAATGGGTCCTTCAGAAATCATGTG 863
Db	358	ATTTTT	TGTCTGTGGTACCTCAGGCT	ATCGTTTGAATGGGTCCTTCAGAAATCATGTG 299
Qy	864	CTTCCT	CTCATCTTATAGTGGCCCTAT	ATGACATCTACCTGAAACAAGATTTATCAATCA 923
Db	298	CTTCCT	CTCATCTTATAGTGGCCCTAT	ATGACATCTACCTGAAACAAGATTTATCAATTA 239
Qy	924	TGTCGT	ACCTTAAGCCCCACAAACAAG	ATACCAATTCCTTCCTTTGTATACAGACAGG 983
Db	238	TGTCAT	ATCTAAGCCCCCGCAACAAAG	ATACCAATTCCTTCCTTTGTATATAGACAGG 179
Qy	984	AGTCTA	GGCAGACTAGGTACTGGCA	TATGGCAGTATCACAACTCTACTCAGTTTCTACTA 1043
Db	178	AGTCTA	GGTGCACCTAGGTACTGGCA	TATGGCAGTATCACAACTCTACTCAGTTTCTACTA 119
Qy	1044	CAAACT	ATCTCAAGAAATAAATGGT	GTGATCGAAACAGGTCACTACCTCCCTGGTCACTT 1103
Db	118	CAAACT	ATCTCAAGAACTAAATGGG	ACATGTGAAACGGGTGCCGACTCCCGTCACTT 59

Qy	864	CTTCCCTCTCATTTCTTAGTGGCCCCCTATGAGCATCTACACTGAAACAAGATTATACAAATCA	922
Db	298	CTTCCCTCTCATTTCTTAGTGGCCCCCTATGAGCATCTACACTGAAACAAGATTATACAAATTA	239
Qy	924	TGTCGTACCTAAGGCCCCACAACAAAAGAGTACCCATTCTTCCCTTTTCTTATCAGAGCAGG	983
Db	238	TGTCATATCTAAGCCCCCGCAACAAGAAGTACCCATTCTTCCCTTTTGTATAGGAGCAGG	179
Qy	984	AGTGCTAGGCAGACTAGGTACTGGCATTTGGCAGTATCACAACTCTCTACTCAGTTTCTACTA	1043
Db	178	AGTGCTAGGTGCATAGTACTGGCATTTGGCGGTATCACAACTCTCTACTCAGTTTCTACTA	119
Qy	1044	CAAACTATCTCAAGAAATAAATGGTGATCATGGAAACAGGTCACTGATCCCTCGTGTCACCTT	1103
Db	118	CAAACTATCTCAAGAACTAAATGGGACATGGAACGGGTGCGCCGACTCCCTGGTCACCTT	59

```

QY 1104 GCAAGATCAACTTAACCTCCCTAGCAGCAGTAGTCCTTCAAAATCGAGAGCTTTAGAC 1161
Db |||||||
58 GCAAGATCAACTTAACCTCCCTAGCAGCAGTAGTCCTTCAAAATCGAGAGCTTTAGAC 1

RESULT 26
BX409304
LOCUS
DEFINITION
  BX409304 Homo sapiens PLACENTA mRNA linear EST 03-MAY-2004
  5-PRIME, mRNA sequence.
ACCESSION
  BX409304
VERSION
  BX409304.2 GI:46956516
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 792)
  Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
  Full-length cDNA libraries and normalization
  Unpublished (2001)
  On May 15, 2003 this sequence version replaced gi:30766967.
  Contact: Genoscope
  Genoscope - Centre National de Sequencage
  BP 191 91006 EVRY cedex - France
  Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
  1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
  end enriched, double-strand cDNA was digested with Not I and cloned
  into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
  was not normalized. Library was constructed by Life Technologies, a
  division of Invitrogen.
  This sequence belongs to sequence cluster 4215.r
  For more information about this cluster, see
  http://www.genoscope.cns.fr/cdna?s=CS08ABE011ZH12_AE01057_1&c=4215.r

FEATURES
  source
    Location/Qualifiers
      1..792
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="CS08E012YJ24"
        /tissue_type="PLACENTA"
        /clone_lib="Homo sapiens PLACENTA"
        /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
        with a NotI-oligo(dT) primer. Five prime end enriched,
        double-strand cDNA was digested with Not I and cloned into
        the Not I and EcoRV sites of the pCMVSPORT 6 vector.
        Library was not normalized."

ORIGIN
  Query Match 37.5%; Score 555.8; DB 5; Length 792;
  Best Local Similarity 88.0%; Pred. No. 5.4e-153;
  Matches 610; Conservative 0; Mismatches 82; Indels 1; Gaps 1;

QY 29 TTACTGTTCTTTACCCCTTTTCGCTCTCACTGACCCCTCATGCTGTGTACAAACA 88
Db |||||||
1 TTACTGTTCTTTACCCCTTTTCGCTCTCACTGACCCCTCATGCTGTGTACAAACA 60

QY 89 GTAGCTCCCTTACCAAGAGTTTCTATGAGACGCGCTTCTCGGAATATTGATGCC 148
Db |||||||
61 GTAGCTCCCTTACCAAGAGTTTCTATGAGAGATGAGCGTCCGGAATATTGATGCC 120

QY 149 CATCATATAGGAGTTTATTAAGGAAATCCACCTTCACTGCCACACCCATATGCC 208
Db |||||||
121 CATCGTATAGGAGTTTCTAAGGGAACCCCACTTCACTGCCACACCCATATGCC 180

QY 209 GCAACTGCTATAACTGTGCACCTCTTTGTCATGATGCAATCAATATTGGACAGGA 268
Db |||||||
181 GCAACTGCTATACTGTGCACCTCTTTGTCATGATGCAATCAATATTGGACAGGA 240

QY 269 AATGATTAATCTAGTTCCTGAGGACTTGGACCACTGTCTTGGACTTACTTCA 328
Db |||||||
241 AATGATTAATCTAGTTCCTGAGGACTTGGAGTCACTGTCTTGGACTTACTTCA 300

```

```

QY 329 CCATATACAGTATGTCTGATGGGGGTGGAAATTCAGGTACGCAAGAAAAACAAGTAA 388
Db |||||||
301 CCAAAACCTGGTATGTCTGATGGGGGTGGAGTTTCAAGATCAGGAGAAAAACATGAA 360

QY 389 AGGAAGCAATCTCCAACTGACCCGGGGACATAGCAGCCCTAGCCCTACAAAGGACTAG 448
Db |||||||
361 AGAAGAAATNTNCCAACTAACCCGGNTATCATGGCACCTNTAGCCCTTCAAAAGGACTAG 420

QY 449 TTCTCTCAAAACTACATGAAGACCTCCGTACCCATACCTCGCCTGGTGAGCTATTTAATA 508
Db |||||||
421 ATCTCTCAAAACTACATGAAGACCTCCGTACCCATACCTCGCCTGGNAAGTCTATTNAATA 480

QY 509 CACCCCTCACTCGGCTCCATGAGGTCTCAGCCCAAAACCCCTACTAATCTGTGGATGTGCC 568
Db |||||||
481 CCACCCCTCACTGNCCTCCATGAGGTTTCGTTCAAAACCCCTACTAATCTGTATATGCC 540

QY 569 TCCCTCTGACATTCAGGCCATACATTTCAATCCCTGCTCTGCAACCAATGGAACAACCTTCA 628
Db |||||||
541 TCCCTCTGACATTCAGGCCATATGCTCAATTCCTGTACTGAACTGCAACACCTTCA 600

QY 629 GCACAGAAATAAACA-CCACTTCCGTTTGTAGTAGGACCTCTTCTTCCAACTCTGGAATA 687
Db |||||||
601 GCACAGTAAATAAATACCCCTTCTGTTTAGTAGGACCTCTTCTTCAATCTGTTAATA 660

QY 688 ACCATACCTCAAACTCACTGCTGTGTAATAATTT 720
Db |||||||
661 TCTCTTTCTCAAACTTACCTGCTGTGTAATAATTT 693

RESULT 27
BX430055/c
LOCUS
DEFINITION
  BX430055 Homo sapiens PLACENTA Homo sapiens cDNA clone CS08E013Y120
  3-PRIME, mRNA sequence.
ACCESSION
  BX430055
VERSION
  BX430055.2 GI:47006728
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 814)
  Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
  Full-length cDNA libraries and normalization
  Unpublished (2001)
  On May 15, 2003 this sequence version replaced gi:30778885.
  Contact: Genoscope
  Genoscope - Centre National de Sequencage
  BP 191 91006 EVRY cedex - France
  Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
  1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
  end enriched, double-strand cDNA was digested with Not I and cloned
  into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
  was not normalized. Library was constructed by Life Technologies, a
  division of Invitrogen.
  This sequence belongs to sequence cluster 4215.r
  For more information about this cluster, see
  http://www.genoscope.cns.fr/cdna?s=CS08AB002ZG11_CS00184_1&c=4215.r

```

FEATURES

source

```

Location/Qualifiers
  1..814
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="CS08E013Y120"
    /tissue_type="PLACENTA"
    /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
    with a NotI-oligo(dT) primer. Five prime end enriched,
    double-strand cDNA was digested with Not I and cloned into
    the Not I and EcoRV sites of the pCMVSPORT 6 vector.
    Library was not normalized."

```

ORIGIN	Query Match Best Local Similarity 36.3%; Score 537.6; DB 5; Length 814; Matches 671; Conservative 0; Mismatches 105; Indels 11; Gaps 6;		ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 586) NIH-MGC http://mgs.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-r@mail.nih.gov Tissue procurement: ATCC CDNA Library Preparation: Ling Hong/Rubin Laboratory DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml Seq primer: -40RP from Gibco High quality sequence stop: 519.
			REFERENCE	1 (bases 1 to 586) NIH-MGC http://mgs.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-r@mail.nih.gov Tissue procurement: ATCC CDNA Library Preparation: Ling Hong/Rubin Laboratory DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml Seq primer: -40RP from Gibco High quality sequence stop: 519.
			AUTHORS	NIH-MGC http://mgs.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-r@mail.nih.gov Tissue procurement: ATCC CDNA Library Preparation: Ling Hong/Rubin Laboratory DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml Seq primer: -40RP from Gibco High quality sequence stop: 519.
			JOURNAL	NIH-MGC http://mgs.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-r@mail.nih.gov Tissue procurement: ATCC CDNA Library Preparation: Ling Hong/Rubin Laboratory DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml Seq primer: -40RP from Gibco High quality sequence stop: 519.
			COMMENT	NIH-MGC http://mgs.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-r@mail.nih.gov Tissue procurement: ATCC CDNA Library Preparation: Ling Hong/Rubin Laboratory DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml Seq primer: -40RP from Gibco High quality sequence stop: 519.
			FEATURES	Location/Qualifiers 1..586 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:2907101" /tissue_type="choriocarcinoma" /lab_host="NIH MGC 21" /note="Organ: placenta; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
			source	Location/Qualifiers 1..586 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:2907101" /tissue_type="choriocarcinoma" /lab_host="NIH MGC 21" /note="Organ: placenta; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
			ORIGIN	Query Match 36.2%; Score 536.4; DB 2; Length 586; Best Local Similarity 94.7%; Pred. No. 2.8e-147; Matches 555; Conservative 0; Mismatches 31; Indels 0; Gaps 0; 703 CTCACCTGTGTAAATTTAGCAATCTATAGACACACAGCTCCCAATGCATCAGTGG 762 1 CTCACCTGTGTAAATTTAGCAATCTATAGACACACAGCTCCCAATGCATCAGTGG 60 763 GTAAACACCTCCACACAGAAATAGTCTGCCTACCCCTCAGGAATATTTTGTCTGTGTACC 822 61 GTAACTCTCTCCACACAAATAGTCTGCCTACCCCTCAGGAATATTTTGTCTGTGTACC 120 823 TCAGCCTATCATTTGTTGAATGCTCTTCAGAACTCTATGTGCTTCTCTCTCTTAGTG 882 121 TCAGCCTATCGTTGTTGAATGCTCTTCAGAACTCTATGTGCTTCTCTCTCTTAGTG 180 883 CCCCTTATGACCATCTACACTGACAAAGATTTTATACAATCATCTCGTACCTAAGCCCCAC 942 181 CCCCTTATGACCATCTACACTGACAAAGATTTTATACAATCATCTCGTACCTAAGCCCCG 240 943 AACAAAGAGTACCCATTCTTCCTTTTGTATCAGACGAGGAGTGTGTAGCAGACTAGGT 1002 241 AACAAAGAGTACCCATTCTTCCTTTTGTATCAGACGAGGAGTGTGTAGCAGACTAGGT 300 1003 ACTGGCATTTGGCAGTATCACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAATA 1062 301 ACTGGCATTTGGCAGTATCACAACCTCTACTCAGTTCTACTACAACTATCTCAAGAATA 360 1063 AATGGTGACATGGAACAGGTCTACTGCTCCCTGCTACCTTGCAGAGATCAACTTAATCTCC 1122 361 AATGGTGACATGGAACAGGTCTACTGCTCCCTGCTACCTTGCAGAGATCAACTTAATCTCC 420 1123 CTAGCAGCAGTACTCTCTTCAAAATCGAAGAGCTTTTATAGCTTGTAAACCCGCAAGAGGG 1182 421 CTAGCAGCAGTACTCTCTTCAAAATCGAAGAGCTTTTATAGCTTGTAAACCCGCAAGAGGG 480

RESULT 28	BE019603	586 bp	mRNA	linear	EST 06-JUN-2000
LOCUS	bas4f03.y1	NIH_MGC_21	Homo sapiens	CDNA clone	IMAGE:2907101 5'
DEFINITION	similar to TR:O95244 O95244 ENVELOPE PROTEIN ; contains Alu repetitive element, mRNA sequence.				
ACCESSION	BE019603				
VERSION	BE019603.1	GI:8279682			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				

/clone.lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dN) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN

Query Match	32.3%	Score 478.6;	DB 5;	Length 760;
Best Local Similarity	83.3%	Pred. No. 4.3e-130;		
Matches 604;	Conservative 0;	Mismatches 114;	Indels 7;	Gaps 5;

Qy	682	GAATAACCCATACCTCAACACCTCACCTGTGTAAATTTAGCAATACT-ATAGACACAAC	740
Db	727	GCACACACCCAGCCCTCCCTCGGGGGGCAAAATCTCGCTATTCTCTCTAATCAAC	668
Qy	741	CAGCTCCCAATGCATCAG---GTGGGTAACACTCCACACAGAAATAGTCTGCCATCCCTC	797
Db	667	CGACCACCCATGGATCAGGGTGAGTCGTCCCTCTCTCACAAAATAGTTTGTCTATACCTCC	608
Qy	798	AGGAATATTTTGTCTGTGTACCTCAGCCTATCATTTTGAATGGCTCTTCAGAATC	857
Db	607	AGAAATAGTGGTTGTGTGAGGCTCCTCAGCCACCGTAGTATGAAGGGCTCTTCAGAAC	548
Qy	858	TATGTGCTTCTCTCAATCTTAGTGCCCTATGACCATCTACACTGAAC-AAAGATTTAT	916
Db	547	ATGTGC-TCTCGCAATCTAAGAGCCCTATGACCATCGACACTGAACTAAGATTTAG	489
Qy	917	ACAAATCATGTGTACCTTAAGCCCAACAAAGAGTACCCATCTTCCTTTTGTATCA	976
Db	488	CCCATATGTCAVATCTAAGCCCAACAAAGAGTCCCATCTTCCTTTTGTATAG	429
Qy	977	GAGCAGGAGTCTAGGCAGACTAGGTACTGGCATTGGCAGTATCACAA-CCTCTACTCAG	1035
Db	428	AAGCAGGAGTCTAGGAGCACTAGGAATGGCATTTGCGTATCAAAAGCTCTACTCAG	369
Qy	1036	TTCTACTACAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTG	1095
Db	368	TTCTACTAGGACTATCTCAAGAACTAAATGGGACATGGAACGGGTGCGCGACTCCCTG	309
Qy	1096	GTCACCTTGAAGATCAACTTAACTCCCTAGCAGCAGTACTCTTCAAAATCGAAGACT	1155
Db	308	GTCACCTTGAAGATCAACTTAACTCCCTAGCAGCAGTACTCTTCAAAATCGAAGACT	249
Qy	1156	TTAGACTTGTCTAACCGCCAAAGAGGGGAACTGTATTATTTTGGAGAGAACGCTGT	1215
Db	248	TTAGACTTGTCTAACCGCTGAAGAGGGGGACCTGTATTATTTTGGGGAAGATGCTGT	189
Qy	1216	TATTATGTTAATCAATCCAGAAATGTCTCAGAAAGTTAAAGAAATTCGAGATCGAATA	1275
Db	188	TACTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATA	129
Qy	1276	CAATGTAGCAGCAGGAGCTTCAAAACACCAACCGCTGGGGCTCCCTCAGCCAATGGATG	1335
Db	128	CAACGTAGCAGCAGGAGCTTCGAAACACTGGACCCCTGGGGCTCCCTCAGCCAATGGATG	69
Qy	1336	CCCTGGTTCCTCCCTTCTTAGGACCTTAGCAGCTCTAATATTGTTACTCCTCTTTGGA	1395
Db	68	CCCTGGATTCCTCCCTTCTTAGGACCTTAGCAGCTAATATTGCTACTCCTCTTTTGA	9
Qy	1396	CCCTG 1400	
Db	8	CCCTG 4	

Search completed: January 22, 2005, 19:31:41
Job time : 4420.62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2005, 20:15:26 ; Search time 114 Seconds

(without alignments)
1551.347 Million cell updates/sec

Title: US-09-319-156b-10

Perfect score: 2634

Sequence: 1 MALPYHTFLFTVLLPPFALT.....KPVSSRIEAVKQLWQLQMEP 493

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2632	99.9	493	2 AAW71068	AAW71068 Multiple
2	2632	99.9	493	2 AAW99553	Aaw99553 Protein e
3	2632	99.9	493	8 ADG14847	Adg14847 MSRV asso
4	2622.5	99.6	541	3 AAB08195	Aab08195 Amino aci
5	2613	99.2	542	3 AAB19069	Aab19069 Amino aci
6	2605	98.9	493	7 ADB84401	Adb84401 MSRV-1 as
7	2447	92.9	542	5 AAE25054	Aae25054 Human EMB
8	2319	88.0	538	2 AAY08622	Aay08622 Human sec
9	2319	88.0	538	3 AAY67313	Aay67313 Human syn
10	2319	88.0	538	5 AAE14540	Aae14540 Human sec
11	2319	88.0	538	7 ADC38777	Adc38777 Human sec
12	2314	87.9	538	4 AAB75138	Aab75138 HERV-W en
13	2314	87.9	538	4 AAB67652	Aab67652 Amino aci
14	2314	87.9	540	3 ABP40980	Abp40980 Human ret
15	2314	87.9	685	3 ABP40988	Abp40988 Human ret
16	2314	87.9	846	3 ABP41050	Abp41050 HERV-7q e
17	2308.5	87.6	544	7 ADE09243	Ade09243 Novel pro
18	2308.5	87.6	544	7 ADF60170	Adf60170 Human con
19	2237	84.9	538	2 AAW99345	Aaw99345 Human end
20	1235.5	46.9	263	4 AAM18008	Aam18008 Peptide #
21	1235.5	46.9	263	4 AAB37043	Aab37043 Peptide #
22	1235.5	46.9	263	4 AAM30519	Aam30519 Peptide #
23	1235.5	46.9	263	4 ABB31812	Abb31812 Peptide #
24	1235.5	46.9	263	4 ABB22358	Abb22358 Protein #
25	1235.5	46.9	263	4 AAM70185	Aam70185 Human bon

26	1235.5	46.9	263	4 AAM57770	Aam57770 Human bra
27	1235.5	46.9	263	4 ABG51885	Abg51885 Human liv
28	1235.5	46.9	263	4 AAM05648	Aam05648 Peptide #
29	1235.5	46.9	263	5 ABG39819	Abg39819 Human pep
30	1130.5	42.9	446	4 ABG07532	Abg07532 Novel hum
31	1130.5	42.9	446	4 ABG20007	Abg20007 Novel hum
32	1122	42.6	531	5 ABP63036	Abp63036 Human pol
33	965.5	36.7	368	4 ABG28311	Abg28311 Novel hum
34	965.5	36.7	633	4 ABG28306	Abg28306 Novel hum
35	591	22.4	120	4 AAB75137	Aab75137 HERV-W tr
36	589	22.4	122	5 ADK35467	Adk35467 Novel hum
37	571.5	21.7	563	7 ADG86762	Adg86762 Chimeric
38	551.5	20.9	567	2 AAR71700	Aar71700 Spleen ne
39	517	19.6	162	7 ADB84404	Adb84404 MSRV-1 as
40	516	19.6	162	2 AAW71069	Aaw71069 Multiple
41	516	19.6	162	2 AAW99554	Aaw99554 Protein e
42	516	19.6	162	8 ADG14850	Adg14850 MSRV asso
43	514.5	19.5	584	4 ABG22871	Abg22871 Novel hum
44	514.5	19.5	584	4 ABG19851	Abg19851 Novel hum
45	514.5	19.5	584	4 ABG05606	Abg05606 Novel hum

ALIGNMENTS

RESULT 1
AAW71068
ID AAW71068 standard; protein; 493 AA.

XX AC AAW71068;

XX DT 29-DEC-1998 (first entry)

XX DE Multiple sclerosis associated retrovirus protein 3.

KW Multiple sclerosis associated retrovirus; MSRV; MS; pol gene; gag gene;
KW env gene; rheumatoid arthritis-associated virus.

OS Multiple sclerosis associated retrovirus.

PH Key Location/Qualifiers

FT Misc-difference 39 /note= "Encoded by tga"

XX PN WO9823755-A1.
XX PD 04-JUN-1998.

XX PF 26-NOV-1997; 97WO-IB001482.
XX PR 26-NOV-1996; 96US-00756429.

XX PA (INMR) BIO MERIEUX.

PI Perron H, Beseme F, Bedin F, Paranhos-Baccala G;
PI Komurian-Pradel F, Jollivet-Reynaud C, Mandrand B;

XX DR WPI; 1998-322732/28.
XX DR N-PSDB; AAV43217.

XX PT New nucleic acid from retroviruses - useful for diagnosis, prevention and
XX treatment of, e.g. multiple sclerosis.

XX PS Disclosure; Page 185-187; 286pp; English.

CC The present sequence represents a multiple sclerosis (MS) associated
CC retrovirus (MSRV) protein used in the method of the invention. The
CC invention provides complete or partial genomic sequences of the MSRV-1
CC pol gene, gag gene and env gene, and polypeptides encoded by these genes.
CC The invention also provides antibodies raised against the polypeptides.
CC The genomic sequences, polypeptides and antibodies are also claimed
CC useful for diagnosing infection by MS and rheumatoid arthritis-associated
CC viruses, and also for prevention and treatment of infection with these

```

CC viruses
XX PA (INMR ) BIO MERIEUX.
SQ Sequence 493 AA;

Query Match 99.9%; Score 2632; DB 2; Length 493;
Best Local Similarity 100.0%; Pred. No. 4.5e-229;
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHTFTLVLLPPFALTAPPCCCTTSSSPYQBFLEXTRLPGNIDAPSYRSLKSGNS 60
Db 1 MALPYHTFTLVLLPPFALTAPPCCCTTSSSPYQBFLEXTRLPGNIDAPSYRSLKSGNS 60
QY 61 TFTAHTMPNRCVNSATLCMHANTHYWTGMINPSCPGGLGATVCWTFYTHTSMSGGGI 120
Db 61 TFTAHTMPNRCVNSATLCMHANTHYWTGMINPSCPGGLGATVCWTFYTHTSMSGGGI 120
QY 121 QGOAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRLTHRLVSLFNTTLRLHEVSA 180
Db 121 QGOAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRLTHRLVSLFNTTLRLHEVSA 180
QY 181 QNPTNCWMCPLPHFRPYISIPVPEQWNNFSTEINTTSLVGLPLVSNLEITHTSNLTCKVF 240
Db 181 QNPTNCWMCPLPHFRPYISIPVPEQWNNFSTEINTTSLVGLPLVSNLEITHTSNLTCKVF 240
QY 241 SNTIDTSSQICRWVTPPTTRIVCLPSGIFVCGTSAVHCLNGSSESMCFSLFVPPMTIY 300
Db 241 SNTIDTSSQICRWVTPPTTRIVCLPSGIFVCGTSAVHCLNGSSESMCFSLFVPPMTIY 300
QY 301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLRGLGTGIGSITTSQFYKLSQEQINGDMEQ 360
Db 301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLRGLGTGIGSITTSQFYKLSQEQINGDMEQ 360
QY 361 VTDSLVTLOQLNSLAAVLQNRALDILLTAKRGCTCLFLGEBRCYVYNOSRIVTEKVK 420
Db 361 VTDSLVTLOQLNSLAAVLQNRALDILLTAKRGCTCLFLGEBRCYVYNOSRIVTEKVK 420
QY 421 IRDRIQCRABELQNTERWGLLSQMPWLPFLGLPLAALILLLLFGPCIFNLLVKFVSSRI 480
Db 421 IRDRIQCRABELQNTERWGLLSQMPWLPFLGLPLAALILLLLFGPCIFNLLVKFVSSRI 480
QY 481 EAVKLQWVLOMEP 493
Db 481 EAVKLQWVLOMEP 493

RESULT 3
ADGI4847
ID ADGI4847 standard; protein; 493 AA.
XX AC ADGI4847;
XX DT 26-FEB-2004 (first entry)
XX DE MSRV associated protein #7.
XX KW pol gene; retrovirus; multiple sclerosis; rheumatoid arthritis.
XX PR

CC viruses
XX PA (INMR ) BIO MERIEUX.
SQ Sequence 493 AA;

Query Match 99.9%; Score 2632; DB 2; Length 493;
Best Local Similarity 100.0%; Pred. No. 4.5e-229;
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHTFTLVLLPPFALTAPPCCCTTSSSPYQBFLEXTRLPGNIDAPSYRSLKSGNS 60
Db 1 MALPYHTFTLVLLPPFALTAPPCCCTTSSSPYQBFLEXTRLPGNIDAPSYRSLKSGNS 60
QY 61 TFTAHTMPNRCVNSATLCMHANTHYWTGMINPSCPGGLGATVCWTFYTHTSMSGGGI 120
Db 61 TFTAHTMPNRCVNSATLCMHANTHYWTGMINPSCPGGLGATVCWTFYTHTSMSGGGI 120
QY 121 QGOAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRLTHRLVSLFNTTLRLHEVSA 180
Db 121 QGOAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRLTHRLVSLFNTTLRLHEVSA 180
QY 181 QNPTNCWMCPLPHFRPYISIPVPEQWNNFSTEINTTSLVGLPLVSNLEITHTSNLTCKVF 240
Db 181 QNPTNCWMCPLPHFRPYISIPVPEQWNNFSTEINTTSLVGLPLVSNLEITHTSNLTCKVF 240
QY 241 SNTIDTSSQICRWVTPPTTRIVCLPSGIFVCGTSAVHCLNGSSESMCFSLFVPPMTIY 300
Db 241 SNTIDTSSQICRWVTPPTTRIVCLPSGIFVCGTSAVHCLNGSSESMCFSLFVPPMTIY 300
QY 301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLRGLGTGIGSITTSQFYKLSQEQINGDMEQ 360
Db 301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLRGLGTGIGSITTSQFYKLSQEQINGDMEQ 360
QY 361 VTDSLVTLOQLNSLAAVLQNRALDILLTAKRGCTCLFLGEBRCYVYNOSRIVTEKVK 420
Db 361 VTDSLVTLOQLNSLAAVLQNRALDILLTAKRGCTCLFLGEBRCYVYNOSRIVTEKVK 420
QY 421 IRDRIQCRABELQNTERWGLLSQMPWLPFLGLPLAALILLLLFGPCIFNLLVKFVSSRI 480
Db 421 IRDRIQCRABELQNTERWGLLSQMPWLPFLGLPLAALILLLLFGPCIFNLLVKFVSSRI 480
QY 481 EAVKLQWVLOMEP 493
Db 481 EAVKLQWVLOMEP 493

RESULT 2
AAW99553
ID AAW99553 standard; protein; 493 AA.
XX AC AAW99553;
XX DT 17-OCT-2003 (revised)
XX DT 27-AUG-2003 (revised)
XX DT 08-JUN-1999 (first entry)
XX DE Protein encoded by clone C15 from MSRV-1.
XX KW Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
XX KW rheumatoid polyarthritis.
XX OS Viruses.
XX FH Key Location/Qualifiers
XX FT Misc-difference 39 /label= unknown
XX FT
XX PN FR2765588-A1.
XX PD 08-JAN-1999.
XX PF 07-JUL-1997; 97FR-00008816.
XX PR 07-JUL-1997; 97FR-00008816.

```

OS Multiple sclerosis associated retrovirus.

XX US2003198647-A1.

XX 23-OCT-2003.

XX 03-APR-2002; 2002US-00114104.

XX 26-NOV-1996; 96US-00756429.

XX 26-NOV-1997; 97US-00979847.

XX (INMR) BIO MERIEUX.

XX Perron H, Beseme F, Bedin F, Paranhos-Baccala G;

PI Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B, Garson JA, Tuke PW;

XX WPI; 2004-032461/03.

XX New isolated nucleic acid and their fragments having the pol gene of a

PT retrovirus, useful for diagnosing, preventing and/or treating multiple

PT sclerosis and/or rheumatoid arthritis.

XX Disclosure; SEQ ID NO 106; 193pp; English.

XX The invention relates to an isolated nucleic acid which comprises the pol

CC Gene of a retrovirus associated with multiple sclerosis or rheumatoid

CC arthritis. The methods and compositions of the present invention are

CC useful for diagnosing, preventing and/or treating multiple sclerosis

CC and/or rheumatoid arthritis. The present sequence is used in the

CC exemplification of the invention.

XX Sequence 493 AA;

Query Match 99.9%; Score 2632; DB 8; Length 493;

Best Local Similarity 100.0%; Pred. No. 4.5e-229;

Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHTFTVLLPPALTAPPPCCCTTSSPYQEFXRTPLPGNIDAPSYRLSKGNS 60

DB 1 MALPYHTFTVLLPPALTAPPPCCCTTSSPYQEFXRTPLPGNIDAPSYRLSKGNS 60

QY 61 TFTAHTMPNCVNSATLCHMHANTHYTGKMINPSCPGGLGATVCTYFTHTSMGCGGI 120

DB 61 TFTAHTMPNCVNSATLCHMHANTHYTGKMINPSCPGGLGATVCTYFTHTSMGCGGI 120

QY 121 QGOAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLRLHEVSA 180

DB 121 QGOAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLRLHEVSA 180

QY 181 QNPTNCWMLPLHFRPYISIPVPEQMNPFSTEINTTSVLGVLVSNLEITHTSNLTCKVF 240

DB 181 QNPTNCWMLPLHFRPYISIPVPEQMNPFSTEINTTSVLGVLVSNLEITHTSNLTCKVF 240

QY 241 SNTIDTSSOCIRVTPTRIVCLPSGIFPVCGTSAHYCLNGSESNCFLSFLVPPMTIY 300

DB 241 SNTIDTSSOCIRVTPTRIVCLPSGIFPVCGTSAHYCLNGSESNCFLSFLVPPMTIY 300

QY 301 TEQDLYNHVVPKHNKVPILPFVIRAGVLGRIGTGITTSQPYKLSQELNGDMEQ 360

DB 301 TEQDLYNHVVPKHNKVPILPFVIRAGVLGRIGTGITTSQPYKLSQELNGDMEQ 360

QY 361 VTDSLVTLOQLNSLAQVQLNRRALDLTAKRGCTCLFLGEERCYYVNGSRIVTEKVK 420

DB 361 VTDSLVTLOQLNSLAQVQLNRRALDLTAKRGCTCLFLGEERCYYVNGSRIVTEKVK 420

QY 421 IRRIQCRABELQNTERWGLLSQMPWVLPPLGFLAALILLPGPCIFNLLVKFVSSRI 480

DB 421 IRRIQCRABELQNTERWGLLSQMPWVLPPLGFLAALILLPGPCIFNLLVKFVSSRI 480

QY 481 EAVKLQWVLMQEP 493

DB 481 EAVKLQWVLMQEP 493

RESULT 4

AAB08195

ID AAB08195 standard; protein; 541 AA.

XX AAB08195;

XX 06-AUG-2003 (revised)

DT 04-DEC-2000 (first entry)

XX Amino acid sequence of the MSRV-1 3' env and LTR regions.

XX MSRV-1; pol region; long terminal repeat; LTR; RUS region; retrovirus.

XX Multiple sclerosis associated retrovirus.

XX Key Location/Qualifiers

FT Peptide 1..27

FT Modified-site /note= "signal peptide"

FT Modified-site 58..60 /note= "N-linked glycosylation site"

FT Modified-site 92..94 /note= "N-linked glycosylation site"

FT Modified-site 213..215 /note= "N-linked glycosylation site"

FT Modified-site 233..235 /note= "N-linked glycosylation site"

FT Modified-site 280..282 /note= "N-linked glycosylation site"

FT Cleavage-site 289..290 /note= "N-linked glycosylation site"

FT Cleavage-site 315..316 /note= "putative cleavage site"

FT Modified-site 408..410 /note= "putative cleavage site"

XX WO200047745-A1.

XX 17-AUG-2000.

XX 15-FEB-2000; 2000WO-IB000159.

XX 15-FEB-1999; 99EP-00420041.

XX (INMR) BIO MERIEUX.

XX Paranhos-Baccala G, Perron H, Komurian-Pradel F;

XX WPI; 2000-506097/45.

XX N-PSDB; AAA63826.

XX Nucleotide fragment of LTR-RUS region from Multiple Sclerosis retrovirus (MSRV) used to detect the presence of MSRV-1 retrovirus in a biological sample.

XX Disclosure; Fig 2; 23pp; English.

XX The present sequence is encoded by a nucleotide sequence corresponding to the 3' env region and long terminal repeat sequences from clone CL6 of Multiple Sclerosis retrovirus (MSRV-1). The specification describes a long terminal repeat (LTR)-RUS region which encodes the expression of a MSRV-1 protein. This is unusual for LTRs, in particular in the RUS region. Probes and antibodies to the MSRV-1 retrovirus protein and encoding polynucleotide sequences are used to detect the presence of MSRV-1 retrovirus in a biological sample. (Updated on 06-AUG-2003 to correct OS field.)

XX Sequence 541 AA;

Query Match 99.6%; Score 2622.5; DB 3; Length 541;

Best Local Similarity 99.8%; Pred. No. 3.7e-228;

Matches 492; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 MALPYHTFTLVLLPPALTAPPCCCTTSSSPYQEFLEXTRLPNGNIDAPSYSLSGNS 60
 DB 1 MALPYHTFTLVLLPPALTAPPCCCTTSSSPYQEFLEXTRLPNGNIDAPSYSLSGNS 59
 QY 61 TFTAHTMPNCVNSATLQWANHANYWTGKMINPSCPGGLGATVCTWYFTHTSMDGGI 120
 DB 60 TFTAHTMPNCVNSATLQWANHANYWTGKMINPSCPGGLGATVCTWYFTHTSMDGGI 119
 QY 121 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRTHTRLVSLFNTTLRLHEVSA 180
 DB 120 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRTHTRLVSLFNTTLRLHEVSA 179
 QY 181 QNPTNCWMLPLHFRPYISIPVPEQWNNFSTEINTTSLVGLVSNLEITHNTSLNLCVKF 240
 DB 180 QNPTNCWMLPLHFRPYISIPVPEQWNNFSTEINTTSLVGLVSNLEITHNTSLNLCVKF 239
 QY 241 SNTIDTSSQICIRWVTPTRIVCLPSGIFPVCGTSAVHCLNGSESCFSLVPPMTIY 300
 DB 240 SNTIDTSSQICIRWVTPTRIVCLPSGIFPVCGTSAVHCLNGSESCFSLVPPMTIY 299
 QY 301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRGTGIGSTTTSTQFYKLSQBEINGDMEQ 360
 DB 300 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRGTGIGSTTTSTQFYKLSQBEINGDMEQ 359
 QY 361 VTDSLVTLDQQLNSLAADVQLQNRALDILLTAKRGGTCLFLGEBRCYVYVNSRIVTEKVK 420
 DB 360 VTDSLVTLDQQLNSLAADVQLQNRALDILLTAKRGGTCLFLGEBRCYVYVNSRIVTEKVK 419
 QY 421 IRDRIQCRABEELQNTERWGLLSQWMPVLPFLGLPAAIILLLLFGPCIFNLLVKFVSSRI 480
 DB 420 IRDRIQCRABEELQNTERWGLLSQWMPVLPFLGLPAAIILLLLFGPCIFNLLVKFVSSRI 479
 QY 481 EAVKLQWVLOMEP 493
 DB 480 EAVKLQWVLOMEP 492

RESULT 5
 AAB19069
 ID AAB19069 standard; protein; 542 AA.
 AC AAB19069;
 XX
 DT 08-FEB-2001 (first entry)
 DE Amino acid sequence of an envelope (env) protein of MSRV-1.
 XX Envelope protein; MSRV-1; superantigen; autoimmune disease; Vbeta16;
 KW Vbeta17; multiple sclerosis; vaccine.
 XX Human endogenous retrovirus.
 OS
 XX WO200057185-A1.
 EN 28-SEP-2000.
 XX
 XX 20-MAR-2000; 2000WO-FR000691.
 XX
 FR 19-MAR-1999; 99FR-00003622.
 FR 28-OCT-1999; 99FR-00013755.
 XX
 XX (INNR) BIO MERIEUX.
 PA
 XX Perron H, Lafont M;
 PI
 XX WPI; 2000-638279/61.
 DR
 DR N-ESDB; AAA96625.
 XX
 XX Detecting superantigen activity, useful for identifying agents for
 PT treatment or prevention of autoimmune disease, from expansion or loss of
 PT particular lymphocyte Vbeta determinants.
 XX
 PS Claim 28; Page 124-126; 134pp; French.

XX The present sequence represents an envelope protein of MSRV-1. The
 CC envelope protein expressed by the endogenous human retrovirus MSRV-1 has
 CC superantigen activity associated with autoimmune disease. The protein can
 CC be detected using the method of the invention. The specification
 CC describes a process for detecting activity of a superantigen in a
 CC biological sample. The process comprises identifying large scale
 CC expansion or loss of lymphocytes that carry at least one of the Vbeta16
 CC and/or Vbeta17 determinants. The method is used to screen for agents that
 CC inhibit the superantigen, especially those associated with MSRV-1 which
 CC is implicated in autoimmune disease, particularly multiple sclerosis.
 CC These agents are potentially useful for treatment or prevention (e.g. as
 CC vaccines) of autoimmune diseases
 XX
 SQ Sequence 542 AA;
 Query Match 99.2%; Score 2613; DB 3; Length 542;
 Best Local Similarity 98.8%; Pred. No. 2.7e-227;
 Matches 487; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MALPYHTFTLVLLPPALTAPPCCCTTSSSPYQEFLEXTRLPNGNIDAPSYSLSGNS 60
 DB 1 MALPYHTFTLVLLPPALTAPPCCCTTSSSPYQEFLEXTRLPNGNIDAPSYSLSGNS 60
 QY 61 TFTAHTMPNCVNSATLQWANHANYWTGKMINPSCPGGLGATVCTWYFTHTSMDGGI 120
 DB 60 TFTAHTMPNCVNSATLQWANHANYWTGKMINPSCPGGLGATVCTWYFTHTSMDGGI 120
 QY 121 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRTHTRLVSLFNTTLRLHEVSA 180
 DB 120 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRTHTRLVSLFNTTLRLHEVSA 180
 QY 181 QNPTNCWMLPLHFRPYISIPVPEQWNNFSTEINTTSLVGLVSNLEITHNTSLNLCVKF 240
 DB 180 QNPTNCWMLPLHFRPYISIPVPEQWNNFSTEINTTSLVGLVSNLEITHNTSLNLCVKF 240
 QY 241 SNTIDTSSQICIRWVTPTRIVCLPSGIFPVCGTSAVHCLNGSESCFSLVPPMTIY 300
 DB 240 SNTIDTSSQICIRWVTPTRIVCLPSGIFPVCGTSAVHCLNGSESCFSLVPPMTIY 300
 QY 301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRGTGIGSTTTSTQFYKLSQBEINGDMEQ 360
 DB 300 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRGTGIGSTTTSTQFYKLSQBEINGDMEQ 360
 QY 361 VTDSLVTLDQQLNSLAADVQLQNRALDILLTAKRGGTCLFLGEBRCYVYVNSRIVTEKVK 420
 DB 360 VTDSLVTLDQQLNSLAADVQLQNRALDILLTAKRGGTCLFLGEBRCYVYVNSRIVTEKVK 420
 QY 421 IRDRIQCRABEELQNTERWGLLSQWMPVLPFLGLPAAIILLLLFGPCIFNLLVKFVSSRI 480
 DB 420 IRDRIQCRABEELQNTERWGLLSQWMPVLPFLGLPAAIILLLLFGPCIFNLLVKFVSSRI 480
 QY 481 EAVKLQWVLOMEP 493
 DB 480 EAVKLQWVLOMEP 493

RESULT 6
 ADB84401
 ID ADB84401 standard; protein; 493 AA.
 AC ADB84401;
 XX
 DT 04-DEC-2003 (first entry)
 DE MSRV-1 associated protein/peptide sequence #10.
 XX multiple sclerosis; rheumatoid arthritis; gag; pol;
 KW reverse transcriptase; ribonuclease H.
 XX Unidentified.
 OS
 XX US2003039664-A1.

XX PD 27-FEB-2003. 181 QNPTNCWCLPLHFRPYISIPVEQWNNFSTEINTTSLVGLVLSNLEITHSTNLCTCVKF 240

XX PF 26-NOV-1997; 97US-00979847. 241 SNTIDTSSQICIRWVTPPTTRIVCLPSGIFPVCTSAHYHCLNGSSESCMCFSLFVLPMTIY 300

XX PR 26-NOV-1996; 96US-00756429. 241 SNTIDTSSQICIRWVTPPTTRIVCLPSGIFPVCTSAHYHCLNGSSESCMCFSLFVLPMTIY 300

XX PA (PERR/) PERRON H. 301 TEODLYNHVVVKPHKRVPLPEVIRAGVLGRGTGIGSTTTTOFYVYKLSOELNGDMEQ 360

XX PA (BESE/) BESEME F. 301 TEQLYNHVVVKPHKRVPLPEVIRAGVLGRGTGIGSTTTTOFYVYKLSOELNGDMEQ 360

XX PA (BEDJ/) BEDIN F. 361 VTDSLVTLODQLNSLAADVQLQNRRLDLLTAKRGGTCLFLGEBRCYVYVNSRIVTEKVK 420

XX PA (PARA/) PARANHOS-BACCALA G. 361 VTDSLVTLODQLNSLAADVQLQNRRLDLLTAKRGGTCLFLGEBRCYVYVNSRIVTEKVK 420

XX PA (KOMA/) KOMURIAN-PRADEL F. 421 IRDRIQCRABELQNTERWGLLSQWMPWLPFLGFLAALILLLLFGPCIFNLLVKFVSSRI 480

XX PA (JOLI/) JOLIVET-REYNAUD C. 421 IRDRIQCRABELQNTERWGLLSQWMPWLPFLGFLAALILLLLFGPCIFNLLVKFVSSRI 480

XX PA (NAND/) MANDRAND B. 481 EAVKLOVQLQMEP 493

XX PA (GARS/) GARSON J A. 481 EAVKLOVQLQMEP 493

XX PA (TUKE/) TUKE P W.

XX PI Perron H, Beseme F, Bedin F, Paranhos-Baccala G; Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B, Garson JA, Tuke PW; WPI; 2003-512253/48.

XX PT New isolated or purified nucleic acid associated with multiple sclerosis and/or rheumatoid arthritis, useful for detecting a virus associated with multiple sclerosis or rheumatoid arthritis in a biological sample.

XX PS Claim 31; Page 78-80; 193pp; English.

XX CC The invention relates to an isolated or purified nucleic acid from a virus associated with multiple sclerosis and/or rheumatoid arthritis, multiple sclerosis-associated virus (MSRV)-1. The nucleic acids comprise pol, gag or reverse transcriptase genes (or their fragments) encoding the proteins or defined peptides (including immunodominant peptides, antigenic peptides or conserved motifs). Also included are a process for detecting a virus associated with multiple sclerosis or rheumatoid arthritis in a biological sample, a nucleic acid probe for the detection of a virus associated with multiple sclerosis or rheumatoid arthritis, a primer for the amplification by polymerisation of a nucleic acid of a viral material associated with multiple sclerosis or rheumatoid arthritis, a polypeptide exhibiting an inhibitory activity on the proteolytic, reverse transcriptase or ribonuclease H activity from MSRV, and an antibody directed against the MSRV-1 virus obtained by immunologically reacting a human or animal body or cells with an immunogenic agent consisting of the antigenic polypeptide defined above. The nucleic acids are useful for detecting a biological sample a virus associated with multiple sclerosis or rheumatoid arthritis, or for detecting in a biological sample, the presence of or exposure to a virus associated with multiple sclerosis or rheumatoid arthritis. The present sequence is a claimed MSRV-associated sequence whose identity cannot be accurately determined. Note: The SEQ ID numbers for the sequences as displayed in the main body of the patent do not match the SEQ ID numbers in the sequence listing. Consequently those sequences mentioned in the claims may not be the sequences the authors intended to claim.

XX SQ Sequence 493 AA;

Query Match 98.9%; Score 2605; DB 7; Length 493;
Best Local Similarity 98.8%; Pred. No. 1.3e-226;
Matches 487; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MALPYHTFLFTVLLPPALAPPCCCTTSSSPQEFLEXTRLPNGNDAPSYSLSGNS 60

DB 1 MALPYHTFLFTVLLPPALAPPCCCTTSSSPQEFLEXTRLPNGNDAPSYSLSGNS 60

QY 61 TFTAHTMPRCYNYSATLCHANYHTWTKMINPSCPGGLGATVCTWYFTHMSDGGI 120

DB 61 TFTAHTMPRCYNYSATLCHANYHTWTKMINPSCPGGLGATVCTWYFTHMSDGGI 120

QY 121 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLTHTRLVLSLFTTLRLHEVSA 180

DB 121 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLTHTRLVLSLFTTLRLHEVSA 180

QY 181 QNPTNCWCLPLHFRPYISIPVEQWNNFSTEINTTSLVGLVLSNLEITHSTNLCTCVKF 240

CC The invention relates to human embryogenesis associated proteins (EMERY)
 CC and nucleic acid molecules encoding such proteins. EMERY sequences are
 CC useful for screening modulators useful for treating or preventing
 CC disorders associated with abnormal expression of EMERY. The disorders
 CC treated include reproductive disorders such as infertility,
 CC endometriosis, endometrial or ovarian tumour; autoimmune/inflammatory
 CC disorder such as acquired immune deficiency syndrome (AIDS), allergies,
 CC contact dermatitis; disorders of the placenta such as preclampsia,
 CC abruptio placentae etc. Sequences of the invention are also useful for
 CC analyzing a proteome of a tissue or a cell type. EMERY proteins are
 CC useful as immunogens for preparing antibodies. Polynucleotides of the
 CC invention are useful for creating knockin humanised animals or transgenic
 CC animals to model human diseases. They are also used in gene therapy. The
 CC present sequence is human EMERY-2 protein

XX
 SQ Sequence 542 AA;

Query Match 92.9%; Score 2447; DB 5; Length 542;
 Best Local Similarity 93.1%; Pred. No. 2.9e-212;
 Matches 459; Conservative 8; Mismatches 26; Indels 0; Gaps 0;

QY 1 MALPYHTFLFTVLLPPFALTAPPCCCTTSSPYQEFELXETRLPGNIDAPSYSLSGNS 60
 DB 1 MALPYCFLFTVLLSPFSLTAPSCCHCTSSPYQAFLEWRMRPHIDAPSYSLSGNP 60
 QY 61 TFTAHTMPNRCYNYSATLCMHANTHYWTGMINPSCPGGLGATVCWTYFTHTSMDSGGI 120
 DB 61 AFTAHTMPNRCYNYSATLCMHANTHYWTGMINPSCPGGLGATVCWTYFTHTSMDSGGV 120
 QY 121 QGOAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRTHTLRLVSLFNTTLRLHEVSA 180
 DB 121 QDOAREKHVKEVISQLTRVHSTSPYKGLDLSKLHETLRTHTLRLVSLFNTTLGLHEVSA 180
 QY 181 QNPTNCWMLPLHFRPYSIPVPEOWNFSTEINTTSLVGLPLVSNLEIHTSNLTCVKF 240
 DB 181 QNPTNCWMLPLHFRPYSIPVPEOWNFSTEINTTSLVGLPLVSNLEIHTSNLTCVKF 240
 QY 241 SNTIDTSSQICIRWTPPTTRIVCLPSGIFVCGTSAHYCLNGSSESCFSLFVPPMTIY 300
 DB 241 SNTIATNSQICIRWTPPTTRIVCLPSGIFVCGTSAHYCLNGSSESCFSLFVPPMTIY 300
 QY 301 TEQDLYNHVVPKPNKRVLPILPFVIRAGVLGRIGTGIGSTTSTQFYKLSQELNGDMEQ 360
 DB 301 TEQDLYNHVVPKPNKRVLPILPFVIRAGVLGRIGTGIGSTTSTQFYKLSQELNGDMEQ 360
 QY 361 VTDSLVTLOQLNSLAAVVLQNRALDILLTAKRGCTCLFLGEECCYVNGSRIVTEKVK 420
 DB 361 VTDSLVTLOQLNSLAAVVLQNRALDILLTAKRGCTCLFLGEECCYVNGSRIVTEKVK 420
 QY 421 IRDRIQCRABEQLONTERWGLLSQMPWVLPGLGPAALILLFPGCIFIPLLKVFSSRI 480
 DB 421 IRDRIQCRABEQLONTERWGLLSQMPWVLPGLGPAALILLFPGCIFIPLLKVFSSRI 480
 QY 481 EAVKLQWVLOMEP 493
 DB 481 EAVKLQWVLOMEP 493

RESULT 8

AAAY08622

ID AAAY08622 standard; protein; 538 AA.

XX AC

AC AAAY08622;

XX XX

DT 10-AUG-1999 (first entry)

XX DE Human secreted protein AJ172_2.

XX XX

Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
 KW bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
 KW cell proliferation; cell differentiation; suppressor; tumour inhibition;
 KW haematopoiesis regulator; activin; inhibin; chemotactic; chemokinetic;
 KW haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;

KW cadherin; tumour invasion suppressor; gene therapy; tissue growth.
 XX Homo sapiens.
 XX WO9926972-A1.
 PN 03-JUN-1999.
 XX 17-NOV-1998; 98WO-US024614.
 PD 21-NOV-1997; 97US-00976110.
 PR 18-MAY-1998; 98US-00080478.
 PR 20-OCT-1998; 98US-00175928.
 XX (GEMY) GENETICS INST INC.
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M;
 XX WPI; 1999-357813/30.
 DR N-PSDB; AAX7526.
 XX New polynucleotides encoding secreted proteins.
 PS Claim 14a; Page 101-103; 142pp; English.

XX This invention describes novel human secreted proteins encoded by
 CC polynucleotides isolated from human adult testes, adult brain, adult
 CC blood or adult placenta, or murine adult bone marrow or thymus cDNA
 CC libraries. The products of the invention are predicted to have biological
 CC activities which would make them suitable for treating, preventing or
 CC ameliorating medical conditions in humans and animals, although no
 CC supporting data is given. Suggested activities include nutritional
 CC activity, cytokine and cell proliferation/differentiation activity,
 CC immune stimulating (e.g. as vaccines) or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, and tumour
 CC inhibition activity. The polynucleotides are also stated to be useful for
 CC gene therapy

XX Sequence 538 AA;

Query Match 88.0%; Score 2319; DB 2; Length 538;
 Best Local Similarity 88.6%; Pred. No. 1.1e-200;
 Matches 436; Conservative 18; Mismatches 38; Indels 0; Gaps 0;

QY 1 MALPYHTFLFTVLLPPFALTAPPCCCTTSSPYQEFELXETRLPGNIDAPSYSLSGNS 60
 DB 1 MALPYHIFLFTVLLSPFSLTAPPCCHCTSSPYQEFLEWRMRPHIDAPSYSLSGTGP 60
 QY 61 TFTAHTMPNRCYNYSATLCMHANTHYWTGMINPSCPGGLGATVCWTYFTHTSMDSGGI 120
 DB 61 TFTAHTMPNRCYNYSATLCMHANTHYWTGMINPSCPGGLGATVCWTYFTHTSMDSGGV 120
 QY 121 QGOAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRTHTLRLVSLFNTTLRLHEVSA 180
 DB 121 QDOAREKHVKEVISQLTRVHSTSPYKGLDLSKLHETLRTHTLRLVSLFNTTLGLHEVSA 180
 QY 181 QNPTNCWMLPLHFRPYSIPVPEOWNFSTEINTTSLVGLPLVSNLEIHTSNLTCVKF 240
 DB 181 QNPTNCWMLPLHFRPYSIPVPEOWNFSTEINTTSLVGLPLVSNLEIHTSNLTCVKF 240
 QY 241 SNTIDTSSQICIRWTPPTTRIVCLPSGIFVCGTSAHYCLNGSSESCFSLFVPPMTIY 300
 DB 241 SNTIATNSQICIRWTPPTTRIVCLPSGIFVCGTSAHYCLNGSSESCFSLFVPPMTIY 300
 QY 301 TEQDLYNHVVPKPNKRVLPILPFVIRAGVLGRIGTGIGSTTSTQFYKLSQELNGDMEQ 360
 DB 301 TEQDLYNHVVPKPNKRVLPILPFVIRAGVLGRIGTGIGSTTSTQFYKLSQELNGDMEQ 360
 QY 361 VTDSLVTLOQLNSLAAVVLQNRALDILLTAKRGCTCLFLGEECCYVNGSRIVTEKVK 420

Db 361 VADSLVTLDQQLNSLAALVQLNRRALDLLTAERGCTCLFGECCYYVNSGIVTEKVE 420
QY 421 IRDRIQRAEELQNTERRWGLLSQMPWVLPFLGFLAALILLFPGCIFNLLKVFSSRI 480
Db 421 IRDRIQRAEELQNTERRWGLLSQMPWVLPFLGFLAALILLFPGCIFNLLKVFSSRI 480
QY 481 EAVKLQWLQME 492
Db 481 EAVKLQWLQME 492
RESULT 9
AAE67313
ID AAY67313 standard; protein; 538 AA.
XX
AC AAY67313;
XX
DT 11-APR-2000 (first entry)
XX
DE Human secreted protein AJ172.2 amino acid sequence.
XX
KW Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;
KW placental pathology; metastasis inhibition; nutritional activity;
KW immune stimulator; haematopoiesis regulator; tissue growth;
KW tumour inhibitor; anti-inflammatory; clone AJ172.2; ATCC_98115.
XX
OS Homo sapiens.
XX
XX WO9960020-A1.
XX
PD 25-NOV-1999.
XX
PF 17-MAY-1999; 99WO-US010915.
XX
PR 18-MAY-1998; 98US-00080478.
PR 20-OCT-1998; 98US-00175928.
XX
XX (GEM) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Mi S, Treacy M;
XX
DR WPI; 2000-116311/10.
DR N-PSDB; AAZ59468.
XX
PT New polynucleotides encoding secreted cDNA libraries, used to develop
PT products for the diagnosis and treatment of neoplastic disease.
XX
PS Claim 15; Page 108-110; 149pp; English.
XX
CC This is the human secreted protein AJ172.2 amino acid sequence, the
CC polynucleotide encoding AJ172.2 was obtained from a human adult testes
CC cDNA library. The invention relates to secreted human and murine
CC proteins. The polynucleotides and proteins are predicted to have
CC biological activities which would make them suitable for treating,
CC preventing or ameliorating medical conditions in humans and animals.
CC Detection of the levels of the proteins can be used for the diagnosis of
CC e.g. pre-eclampsia, placental pathology or cancer. Agents which modulate
CC the expression or function of the proteins may be used for treating a
CC neoplastic disease and inhibiting metastasis. Other suggested activities
CC include nutritional activity (e.g. in feeds), cytokine and cell
CC proliferation/differentiation activity, immune stimulating (e.g. as
CC vaccines) or suppressing activity, haematopoiesis regulating activity,
CC tissue growth activity, activin/inhibin activity.
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity. The
CC polynucleotide sequences are also stated to be useful for gene therapy
XX
SQ Sequence 538 AA;

Query Match

88.0%; Score 2319; DB 3; Length 538;

Best Local Similarity 88.6%; Pred. No. 1.1e-200;
Matches 436; Conservative 18; Mismatches 38; Indels 0; Gaps 0;
QY 1 MALPYHTFTVLLPFPALTAPPCCCTTSSSSPYQBFLXRTLPNGNIDAPSYLSKGN 60
Db 1 MALPYHTFTVLLPFPALTAPPCCCTTSSSSPYQBFLXRTLPNGNIDAPSYLSKGN 60
QY 61 TPTAHTMPCNCYSATLCHNANTHYWTGKMINPSCPGGLGATVCTYFTHTSMDGGI 120
Db 61 TPTAHTMPCNCYSATLCHNANTHYWTGKMINPSCPGGLGATVCTYFTHTSMDGGI 120
QY 121 QGQAREKQVKEAISQLTRGHSTPSYKGLVLSKHLHETLRTHTRLVSLFNTTLRLHEVSA 180
Db 121 QGQAREKQVKEAISQLTRGHSTPSYKGLVLSKHLHETLRTHTRLVSLFNTTLRLHEVSA 180
QY 181 QNPTNCMCLPLHFRPYISIPVPEQNNFSTEINTTSLVGLVPLVSNLEITHNTNLTCVKF 240
Db 181 QNPTNCMCLPLHFRPYISIPVPEQNNFSTEINTTSLVGLVPLVSNLEITHNTNLTCVKF 240
QY 241 SNTIDTSSQICRWVTPTRIIVCLPSGIFPVCGTSAYHCLNGSSEMCFLSFLVPPMTIY 300
Db 241 SNTIDTSSQICRWVTPTRIIVCLPSGIFPVCGTSAYHCLNGSSEMCFLSFLVPPMTIY 300
QY 301 TEODLYNHVVPKPNKRVLPILPVIRAGVLGRGTGIGSTTSTQFYKLSQBLNGDMEQ 360
Db 301 TEODLYNHVVPKPNKRVLPILPVIRAGVLGRGTGIGSTTSTQFYKLSQBLNGDMEQ 360
QY 361 VTDSLVTLDQQLNSLAALVQLNRRALDLLTAERGCTCLFGECCYYVNSGIVTEKVE 420
Db 361 VTDSLVTLDQQLNSLAALVQLNRRALDLLTAERGCTCLFGECCYYVNSGIVTEKVE 420
QY 421 IRDRIQRAEELQNTERRWGLLSQMPWVLPFLGFLAALILLFPGCIFNLLKVFSSRI 480
Db 421 IRDRIQRAEELQNTERRWGLLSQMPWVLPFLGFLAALILLFPGCIFNLLKVFSSRI 480
QY 481 EAVKLQWLQME 492
Db 481 EAVKLQWLQME 492
RESULT 10
AAE14540
ID AAE14540 standard; protein; 538 AA.
XX
AC AAE14540;
XX
DT 07-MAY-2002 (first entry)
XX
DE Human syncytin protein.
XX
KW Human; syncytin; pre-eclampsia; gestational trophoblast disorder;
KW choriocarcinoma; hydatiform mole; placental site tumour; abortion;
KW envelope gene; human endogenous defective retrovirus; HERV-W.
XX
OS Homo sapiens.
XX
PN WO200204678-A2.
PD 17-JAN-2002.
XX
PF 09-JUL-2001; 2001WO-US021719.
XX
PR 07-JUL-2000; 2000US-0216657P.
XX
PA (GEM) GENETICS INST INC.
XX
PI Keith JC, McCoy JM, Mi S;
XX WPI; 2002-171727/22.
DR N-PSDB; AAD24195.
XX
PT Identifying a compound for treating a subject with or at risk of
PT developing pre-eclampsia, comprises determining whether the expression or

PT activity of syncytin in the cell is modulated in the presence of a test compound.

XX Disclosure; Page 42-43; 43pp; English.

XX The invention relates to identifying compounds which are modulators of syncytin expression. The syncytin modulators are useful in diagnosis and treatment of pre-eclampsia and gestational trophoblast disorders (e.g. choriocarcinoma, hydatiform mole, placental site tumour and missed/incomplete abortion). Syncytin is a human gene derived from the envelope gene of human endogenous defective retrovirus, HERV-W. The present invention is based partly on the discovery that syncytin expression is dramatically reduced in pre-eclampsia, and is also mis-localised to the apical syncytiotrophoblast membrane. The present sequence is human syncytin protein

XX Sequence 538 AA;

Query Match 88.0%; Score 2319; DB 5; Length 538;
Best Local Similarity 88.6%; Pred. No. 1.1e-200;
Matches 436; Conservative 18; Mismatches 38; Indels 0; Gaps 0;

Qy 1 MALPYHFLFTVLLPFPALFAPPCCTTSSPYQBELXTRPLGNIDAPSYSLSKGNS 60
Db 1 MALPYHFLFTVLLPFPALFAPPCCTTSSPYQBELXTRPLGNIDAPSYSLSKGTP 60

Qy 61 TFTAHTMPNRYNSATLCMHANTHYWTGKINPSCPGILGATVCTWYFTHTSMDSGGI 120
Db 61 TFTAHTMPNRYNSATLCMHANTHYWTGKINPSCPGILGATVCTWYFTHTSMDSGGV 120

Qy 121 QGQAREKQVKEAISQLTRGHSTPSYKGLVLSKLHETLTHRLVSLFNTTLRLHEVSA 180
Db 121 QGQAREKQVKEAISQLTRGHSTPSYKGLVLSKLHETLTHRLVSLFNTTLRLHEVSA 180

Qy 181 QNPNCWMLPLHFRPYISIPVQWNNFSTEINTSVLGVPLVSLNLEITHNLICVKF 240
Db 181 QNPNCWMLPLHFRPYISIPVQWNNFSTEINTSVLGVPLVSLNLEITHNLICVKF 240

Qy 241 SNTIDTSSQIRWVTPPTTRIVCLPSGIFVCGTSAVHCLNGSSEMCFLSFLVPPMTIY 300
Db 241 SNTIDTSSQIRWVTPPTTRIVCLPSGIFVCGTSAVHCLNGSSEMCFLSFLVPPMTIY 300

Qy 301 TEQDLYNVKPKHKKVPILPFVIRAGVLRIGTGIGSTTSTQFYKLSQELNGDMEQ 360
Db 301 TEQDLYNVKPKHKKVPILPFVIRAGVLRIGTGIGSTTSTQFYKLSQELNGDMEQ 360

Qy 361 VTDSLVTLOQLNSLAVALQNRALDILLAKRGCTCLFLGEERCYYVNSRIVTEKVK 420
Db 361 VTDSLVTLOQLNSLAVALQNRALDILLAKRGCTCLFLGEERCYYVNSRIVTEKVK 420

Qy 421 IRDRIQCRABELQNTERRGLLSQWMPVLPFLGPLAAILLLFGPCIFNLLVKFVSSRI 480
Db 421 IRDRIQCRABELQNTERRGLLSQWMPVLPFLGPLAAILLLFGPCIFNLLVKFVSSRI 480

Qy 481 EAVKLQWVLMQNE 492
Db 481 EAVKLQWVLMQNE 492

RESULT 11
ADC38777
ID ADC38777 standard; protein; 538 AA.
XX AC ADC38777;
XX AC ADC38777;
XX 18-DEC-2003 (first entry)
XX DE Human secreted protein #63.
XX DE immune disorder; severe combined immunodeficiency; SCID;
XX DE autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
XX DE rheumatoid arthritis; allergic reaction; asthma; myeloid cell deficiency;
XX DE lymphoid cell deficiency; osteoporosis; osteoarthritis;

KW peripheral nervous system disease; peripheral neuropathy;
KW Alzheimer's disease; Parkinson's disease; coagulation disorder;
KW inflammatory disease; systemic inflammatory response syndrome; SIRS;
KW ischaemia-reperfusion injury; Crohn's disease; anaphylaxis;
KW hypersensitivity; regeneration; neural cell proliferation; fertility;
KW tumour; chemokine; human; secreted protein.

XX Homo sapiens.

OS US2002193567-A1.

XX 19-DEC-2002.

XX 02-APR-2002; 2002US-00114893.

XX 11-AUG-1995; 95US-00514014.

PR 05-APR-1996; 96US-00628364.

PR 19-APR-1996; 96US-00635311.

PR 07-JUN-1996; 96US-00659224.

PR 19-JUN-1996; 96US-00664596.

PR 09-JUL-1996; 96US-00677231.

PR 26-JUL-1996; 96US-00686878.

PR 23-AUG-1996; 96US-00701819.

PR 27-SEP-1996; 96US-00721488.

PR 27-SEP-1996; 96US-00721798.

PR 27-SEP-1996; 96US-00721923.

PR 27-SEP-1996; 96US-00721926.

PR 25-OCT-1996; 96US-00738367.

PR 30-OCT-1996; 96US-00739775.

PR 13-JAN-1997; 97US-00783395.

PR 10-APR-1997; 97US-00833823.

PR 02-JUN-1997; 97US-00867677.

PR 05-SEP-1997; 97US-00924838.

PR 06-OCT-1999; 99US-00413232.

XX (GEMY) GENETICS INST INC.

XX Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Bowman MR, Spaulding V, Carlin-Duckett M;
PI Kelleher K;
XX WPI: 2003-657236/62.
XX N-PSDB; ADC38776.

XX Proteins A23021 encoded by clone A23021 from human adult colon, and
PT BDI2716 encoded by clone BDI2716 from human fetal kidney cDNA library,
PT useful for treating e.g. multiple sclerosis and rheumatoid arthritis.

XX Disclosure; SEQ ID NO 135; 412pp; English.

XX The invention relates to a protein comprising fully defined A2302 1
CC protein or BDI27 1 6 protein. The polynucleotides are useful for
CC expressing recombinant proteins for analysis and are also useful as
CC chromosome markers or tags to identify chromosomes or to map related gene
CC positions. The proteins are useful as amino acid supplement, carbon
CC source, nitrogen source and carbohydrate source. The proteins are useful
CC for treating various immune deficiencies and disorders (e.g. severe
CC combined immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
CC sclerosis, systemic lupus erythematosus, rheumatoid arthritis), allergic
CC reactions (e.g. asthma), myeloid or lymphoid cell deficiencies,
CC osteoporosis or osteoarthritis, peripheral nervous system diseases (e.g.
CC peripheral neuropathy, Alzheimer's disease, Parkinson's disease),
CC coagulation disorders, inflammatory diseases (e.g. systemic inflammatory
CC response syndrome (SIRS)), ischaemia-reperfusion injury, Crohn's disease,
CC anaphylaxis and hypersensitivity. Proteins are also useful for inducing
CC tumour immunity, for inducing bone, cartilage, tendon, ligament and/or
CC nerve growth or regeneration, for proliferating neural cells and for
CC regenerating nerve and brain tissue, for inducing fertility and for
CC inhibiting tumour growth. Proteins are also useful as chemokine for
CC mammalian cells (e.g., monocytes, fibroblasts, neutrophils), and also
CC useful as inhibitors of receptor/ligand interactions. The present
CC sequence represents the amino acid sequence of a human secreted protein.

```
SQ Sequence 538 AA;
Query Match      88.0%; Score 2319; DB 7; Length 538;
Best Local Similarity 88.4%; Pred. No. 1.1e-200;
Matches 436; Conservative 19; Mismatches 38; Indels 0; Gaps 0;

QY 1 MALPYHTFLTPVLLPPFALTAPPCCCTTSSSPYQEFLLXRLPGNIDAPSYSLSGNS 60
DB 1 MALPYHIFLTPVLLPSFTLTAPPCCRCMTSSPYQEFLLWRMORPGNIDAPSYSLSGKTP 60
QY 61 TFTAHTMPNPNVNSATLCHWANTHYWTGKWINPSCPGGLGATVCTWYFTHTSMDSGGGI 120
DB 61 TFTAHTMPNPNVNSATLCHWANTHYWTGKWINPSCPGGLGATVCTWYFTHTSMDSGGV 120
QY 121 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRTHTLRLVSLFNTTLRLHEVSA 180
DB 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKLHETLRTHTLRLVSLFNTTLGLHEVSA 180
QY 181 QNPTNCMCLPLHFRPVISIPVPEQNNFSTEINTSVLGPVLSNLEIHTSNLTCVKF 240
DB 181 QNPTNCMCLPLHFRPVISIPVPEQNNFSTEINTSVLGPVLSNLEIHTSNLTCVKF 240
QY 241 SNTIDTSSQICIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSECMCFSLFVPPMTIY 300
DB 241 SNTIYTNISQICIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSECMCFSLFVPPMTIY 300
QY 301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRGTGIGSTTTSTQFYKLSQELNGDMEQ 360
DB 301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRGTGIGSTTTSTQFYKLSQELNGDMEQ 360
QY 361 VTDLSVLTQDQLNSLAAVLQNRALDLLTAKRGGTCLFLGEBRCYVYNOSRIVTEKVK 420
DB 361 VADSLVTQDQLNSLAAVLQNRALDLLTAKRGGTCLFLGEBRCYVYNOSRIVTEKVK 420
QY 421 IRDRIQRAELQNTWEGHLSQWMPVILFGLPLAAILLLFGPCIFNLLYKFSVSSRI 480
DB 421 IRDRIQRAELQNTWEGHLSQWMPVILFGLPLAAILLLFGPCIFNLLYKFSVSSRI 480
QY 481 EAVKLQWVQME 492
DB 481 EAVKLQWEPKMQ 492

RESULT 12
AAB75138
ID AAB75138 standard; protein; 538 AA.
XX
AC AAB75138;
XX
DT 08-AUG-2001 (first entry)
XX
DE HERV-W envelope protein G.
XX
KW Human endogenous retrovirus; HERV-W; HERV; chromosome 7; env protein;
KW envelope protein; multiple sclerosis-related superantigen; vaccine;
KW surface antigen; transmembrane; multiple sclerosis; neuroprotective;
KW antisense-therapy; autoimmune disorder.
XX
OS Human endogenous retrovirus.
XX
FH Key Location/Qualifiers
FT Region 271..317
FT /label= SU
FT /note= "surface protein"
FT Cleavage-site 317..318
FT /note= "SU-TM putative cleavage site"
FT Region 318..538
FT /label= TM
FT /note= "transmembrane subunit"
XX
FN WO200131021-A1.
XX
PD 03-MAY-2001.
```

```
XX
PF 30-OCT-2000; 2000WO-EP010659.
XX
PR 28-OCT-1999; 99BP-00402690.
XX
PA (UYGE-) UNIV GENEVE.
XX
PI Conrad B, Mach B;
XX
DR WPI; 2001-316336/33.
DR N-PSDB; AAH20070.
XX
PT New human retrovirus HERV-W ENV proteins/peptides having superantigen
PT activity useful for diagnosing and treating multiple sclerosis.
XX
PS Claim 1; Fig 7; 94pp; English.
XX
CC On the basis of the PBS t-RNA motif used for the classification of human
CC endogenous retrovirus (HERVs) the full length endogenous provirus which
CC was been located on the long arm of human chromosome 7 (7q21-22) has been
CC designated HERV-W. The present invention describes proteins or peptides
CC (I) having superantigen (SAG) activity comprising the ENV protein (ENV)
CC of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)
CC have neuroprotective activity, and can be used in: vaccines; antisense-
CC therapy; and HERV-W SAG activity-inhibitors. (I) and encoding DNA/RNA are
CC useful for diagnosing multiple sclerosis (MS) or HERV-W-associated
CC disorders. (I) are also useful for identifying substances (and optionally
CC recovering) capable of binding to a retroviral superantigen associated
CC with MS, substances capable of blocking SAG activity and substances
CC capable of blocking transcription or translation of HERV-W retroviral
CC superantigen. A protein or peptide derived from (I), modified to be
CC devoid of SAG activity and being capable of generating an immune response
CC against HERV-W retroviral SAG is useful in therapy. Nucleic acid
CC molecules encoding (I) are useful as vaccines against MS. Substances
CC capable of blocking SAG activity, capable of binding to a retroviral
CC superantigen associated with MS, or capable of blocking transcription or
CC translation of HERV-W retroviral superantigen for use in treating or
CC preventing MS, obtained using (I) are useful for the treatment and
CC prevention of MS. (I) and nucleic acids encoding them are useful for
CC diagnosing autoimmune disease. The present sequence represents the
CC specifically claimed envelope protein of HERV-W designated G
XX
SQ Sequence 538 AA;
Query Match      87.9%; Score 2314; DB 4; Length 538;
Best Local Similarity 88.4%; Pred. No. 3.2e-200;
Matches 435; Conservative 19; Mismatches 38; Indels 0; Gaps 0;

QY 1 MALPYHTFLTPVLLPPFALTAPPCCCTTSSSPYQEFLLXRLPGNIDAPSYSLSGNS 60
DB 1 MALPYHIFLTPVLLPSFTLTAPPCCRCMTSSPYQEFLLWRMORPGNIDAPSYSLSGKTP 60
QY 61 TFTAHTMPNPNVNSATLCHWANTHYWTGKWINPSCPGGLGATVCTWYFTHTSMDSGGGI 120
DB 61 TFTAHTMPNPNVNSATLCHWANTHYWTGKWINPSCPGGLGATVCTWYFTHTSMDSGGV 120
QY 121 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRTHTLRLVSLFNTTLRLHEVSA 180
DB 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKLHETLRTHTLRLVSLFNTTLGLHEVSA 180
QY 181 QNPTNCMCLPLHFRPVISIPVPEQNNFSTEINTSVLGPVLSNLEIHTSNLTCVKF 240
DB 181 QNPTNCMCLPLHFRPVISIPVPEQNNFSTEINTSVLGPVLSNLEIHTSNLTCVKF 240
QY 241 SNTIDTSSQICIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSECMCFSLFVPPMTIY 300
DB 241 SNTIYTNISQICIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSECMCFSLFVPPMTIY 300
QY 301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRGTGIGSTTTSTQFYKLSQELNGDMEQ 360
DB 301 TEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRGTGIGSTTTSTQFYKLSQELNGDMEQ 360
QY 361 VTDLSVLTQDQLNSLAAVLQNRALDLLTAKRGGTCLFLGEBRCYVYNOSRIVTEKVK 420
```

```
Db 361 VADSLVTLQDQLNSLAADVQLNRRALDLTAERGGTCLFLGEECCYYVNSQSGIVTEKVKE 420
QY 421 IRDRIQRAEELQNTERRGGLSQWMPWVLPFLGFLAALILLFLFGPCIFNLLVKFVSSRI 480
Db 421 IRDRIQRAEELQNTERRGGLSQWMPWVLPFLGFLAALILLFLFGPCIFNLLVKFVSSRI 480
QY 481 EAVKLQWLQME 492
Db 481 EAVKLQMEPKMQ 492

RESULT 13
AAB67652
ID AAB67652 standard; protein; 538 AA.
XX
AC AAB67652;
XX
XX 29-MAY-2001 (first entry)
XX
DE Amino acid sequence of a human endogenous retrovirus envelope protein.
DE
KW Envelope protein; HERV; syncytia formation; placental development;
KW syncytia; cancer; cell adhesion.
XX
OS Human endogenous retrovirus.
XX
XX WO200116171-A1.
PN
PD 08-MAR-2001.
XX
XX 01-SEP-2000; 2000WO-FR002429.
XX
XX 01-SEP-1999; 99FR-00011141.
PR 15-SEP-1999; 99FR-00011793.
XX
XX (INRM ) BIO MERIEUX.
PA
PA (INRM ) INST NAT SANTE & RECH MEDICALE.
XX
XX Mallet F, Cosset F, Blond J, Lavillette D, Bouton O, Ruggieri A;
XX
DR WPI; 2001-226676/23.
DR N-PSDB; AAF55630.
XX
XX Detecting expression of human endogenous retrovirus envelope protein in
XX cells of a tissue or culture, from its ability to induce syncytia.
XX
XX Claim 1; Page 41-44; 57pp; French.
XX
XX The present sequence represents a human endogenous retrovirus envelope
XX protein. The specification describes a method for detecting expression of
XX an envelope protein from a human endogenous retrovirus (HERV), in cells,
XX of a tissue or culture. The method comprises detecting syncytia formation
XX due to the fusogenic properties of the envelope protein. Envelope
XX polypeptides and polynucleotides are used to produce therapeutic or
XX prophylactic compositions, particularly for treatment of cancer, to
XX correct defects in placental development (for other natural formation of
XX other types of syncytia), and to promote adhesion of cells in grafts or
XX cellular repair processes. Expression of sequences antisense to the
XX polynucleotide are used to prevent formation of syncytia
XX
XX Sequence 538 AA;
XX
Query Match 87.9%; Score 2314; DB 4; Length 538;
Best Local Similarity 88.4%; Pred. No. 3.2e-200;
Matches 435; Conservative 19; Mismatches 38; Indels 0; Gaps 0;
QY 1 MALPYHTFLFTVLLPPLAPPCCCTSSPYQEFLLXRLPLGNIDAPSYSLKSGNS 60
Db 1 MALPYHTFLFTVLLPPLAPPCCCTSSPYQEFLLXRLPLGNIDAPSYSLKSGTP 60
QY 61 TFTAHTHMPRNCVHSATLCMHANTHYWTGKMINPSCPGGLGVTVCTWYFTQTGMSDGGV 120
```

```
Db 61 TFTAHTHMPRNCVHSATLCMHANTHYWTGKMINPSCPGGLGVTVCTWYFTQTGMSDGGV 120
QY 121 QGQAREKQVKEAISQLTRGHSTSPSPYKGLVSLKSLHETLRLTHRLVSLFNTTLRLHEVSA 180
Db 121 QGQAREKHVKEVISQLTRVHGTSPPYKGLDLSKGLHETLRLTHRLVSLFNTTLRLHEVSA 180
QY 181 QNPNTCWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGLPVLSNLEIHTHTSNLTCVKF 240
Db 181 QNPNTCWMCLPLNFRPYISIPVPEQWNNFSTEINTTSVLVGLPVLSNLEIHTHTSNLTCVKF 240
QY 241 SNTIDTSSOCIRWVTPTRIIVCLPSGIFVCGTSAYHCLNGSSSEMCFLSFLVPPMTIY 300
Db 241 SNTITYTNSQCIRWVTPTRIIVCLPSGIFVCGTSAYHCLNGSSSEMCFLSFLVPPMTIY 300
QY 301 TEODLYNHVVPKPHNKRVPILPFVIRAGVLGRGTGIGSTTSTQFYKLSQBIINGDMEQ 360
Db 301 TEODLYSYVVISKPRNKRVPILPFVIRAGVLGRGTGIGSTTSTQFYKLSQBIINGDMEQ 360
QY 361 VTDSLVTLODQLNSLAADVQLNRRALDLTAERGGTCLFLGEECCYYVNSQSGIVTEKVKE 420
Db 361 VADSLVTLQDQLNSLAADVQLNRRALDLTAERGGTCLFLGEECCYYVNSQSGIVTEKVKE 420
QY 421 IRDRIQRAEELQNTERRGGLSQWMPWVLPFLGFLAALILLFLFGPCIFNLLVKFVSSRI 480
Db 421 IRDRIQRAEELQNTERRGGLSQWMPWVLPFLGFLAALILLFLFGPCIFNLLVKFVSSRI 480
QY 481 EAVKLQWLQME 492
Db 481 EAVKLQMEPKMQ 492

RESULT 14
AAB40980
ID AAB40980 standard; protein; 540 AA.
XX
AC AAB40980;
XX
XX 01-AUG-2002 (first entry)
XX
DE Human retroviral HERV-7q env peptide #3.
XX
XX Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
XX multiple sclerosis.
XX
XX Human endogenous retrovirus.
XX
XX WO9967395-A1.
XX
XX 29-DEC-1999.
XX
XX 23-JUN-1999; 99WO-FR001513.
XX
XX 23-JUN-1998; 98FR-00007920.
XX
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
XX Alliel PM, Perin J, Rieger F;
XX
XX WPI; 2000-160587/14.
XX
XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
XX for diagnosis, treatment and prevention of autoimmune and neurological
XX diseases.
XX
XX Claim 22; Fig 4; 225pp; French.
XX
XX The present invention relates to new nucleic acid sequences of human
XX endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
XX Regulatory elements associated with HERV-7q may alter expression of other
XX genes (even remote genes) on the same chromosome, inducing immunological
XX and/or neurological changes (which may be pathological or protective/
XX curative). HERV-7q peptides can be used to improve efficiency of the
XX immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
```

CC sequences can be used in immunogenic or vaccinating compositions, for
 CC protection against autoimmune diseases, particularly multiple sclerosis.
 CC The peptides may also be used (by sequence comparison) to detect/identify
 CC endogenous retroviruses that are abnormally expressed in cancer,
 CC neuropathologies or other autoimmune diseases. The present sequence was
 CC used to illustrate the invention
 XX
 SQ Sequence 540 AA;

Query Match 87.9%; Score 2314; DB 3; Length 540;
 Best Local Similarity 88.4%; Pred. No. 3.2e-200;
 Matches 435; Conservative 19; Mismatches 38; Indels 0; Gaps 0;

Qy 1 MALPYHTFLFTVLLPFPALTAPPCCCTTSSSPYQEFRLXRLPQNTDAPSYSLSGNS 60
 Db 3 MALPYHIFLFTVLLPFPALTAPPCCRCMTSSSPYQEFRLXRLPQNTDAPSYSLSGTP 62
 Qy 61 TFTAHTMPRNCVNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMDSGGI 120
 Db 63 TFTAHTMPRNCVNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMDSGGV 122
 Qy 121 QGQAREKHVKEAISQLTRGHSTSPYKGLVLSKLHETRLTHRLVSLFNTTLRLHEVSA 180
 Db 123 QGQAREKHVKEIVISQLTRVHGTSSPYKGLDLSKLHETRLTHRLVSLFNTTLRLHEVSA 182
 Qy 181 QNPNCWMLPLHFRPYISIPVPEQWNNFSTEINTSVLVGPLVSNLEITHTSNLTCKVF 240
 Db 183 QNPNCWMLPLHFRPYISIPVPEQWNNFSTEINTSVLVGPLVSNLEITHTSNLTCKVF 242
 Qy 241 SNTIDTSSQICIRWVTPTRIVCLPSGIFPVCGTSAYHCLNGSSESCFLSFLVPPMTIY 300
 Db 243 SNTITYTNSQICIRWVTPPTQIVCLPSGIFPVCGTSAYHCLNGSSESCFLSFLVPPMTIY 302
 Qy 301 TEQDLYNHVVPKPKHNRKVPILPFVIRAGVLGRGTGIGSTTSTQFYKLSQELNGDMEQ 360
 Db 303 TEQDLYSVISKPRKRVPIPLPFVIRAGVLGRGTGIGSTTSTQFYKLSQELNGDMEQ 362
 Qy 361 VTDSLVTLDQQLNSLAAVLQNRALDILLTAERGCTCLFLGEECCYYVNSGIVTEKVK 420
 Db 363 VADSLVTLDQQLNSLAAVLQNRALDILLTAERGCTCLFLGEECCYYVNSGIVTEKVK 422
 Qy 421 IRDRIQRAEELONTERWGLLSQWMPWLPFLGPLAAILILLFGPCIFNLLVKFVSRI 480
 Db 423 IRDRIQRAEELONTERWGLLSQWMPWLPFLGPLAAILILLFGPCIFNLLVKFVSRI 482
 Qy 481 EAVKLQWLVOME 492
 Db 483 EAVKLQWLVOME 494

RESULT 15
 ABP40988
 ID ABP40988 standard; protein; 685 AA.
 XX
 AC ABP40988;
 XX
 DT 01-AUG-2002 (first entry)
 XX
 DE Human retroviral env ORF protein.
 XX
 KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
 KW multiple sclerosis.
 XX
 OS Human endogenous retrovirus.
 XX
 FN WO9967395-A1.
 XX
 PD 29-DEC-1999.
 XX
 PF 23-JUN-1999; 99WO-FR001513.
 XX
 PR 23-JUN-1998; 98FR-00007920.
 XX

(INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Alliel PM, Perin J, Rieger F;

XX WPI; 2000-160587/14.

XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
 PT for diagnosis, treatment and prevention of autoimmune and neurological
 PT diseases.

XX Claim 22; Page 152-154; 225pp; French.

XX The present invention relates to new nucleic acid sequences of human
 CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
 CC Regulatory elements associated with HERV-7q may alter expression of other
 CC genes (even remote genes) on the same chromosome, inducing immunological
 CC and/or neurological changes (which may be pathological or protective/
 CC curative). HERV-7q peptides can be used to improve efficiency of the
 CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
 CC sequences can be used in immunogenic or vaccinating compositions, for
 CC protection against autoimmune diseases, particularly multiple sclerosis.
 CC The peptides may also be used (by sequence comparison) to detect/identify
 CC endogenous retroviruses that are abnormally expressed in cancer,
 CC neuropathologies or other autoimmune diseases. The present sequence was
 CC used to illustrate the invention

XX Sequence 685 AA;

Query Match 87.9%; Score 2314; DB 3; Length 685;

Best Local Similarity 88.4%; Pred. No. 4.5e-200;

Matches 435; Conservative 19; Mismatches 38; Indels 0; Gaps 0;

Qy 1 MALPYHTFLFTVLLPFPALTAPPCCCTTSSSPYQEFRLXRLPQNTDAPSYSLSGNS 60
 Db 131 MALPYHIFLFTVLLPFPALTAPPCCRCMTSSSPYQEFRLXRLPQNTDAPSYSLSGTP 190
 Qy 61 TFTAHTMPRNCVNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMDSGGI 120
 Db 191 TFTAHTMPRNCVNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMDSGGV 250
 Qy 121 QGQAREKHVKEAISQLTRGHSTSPYKGLVLSKLHETRLTHRLVSLFNTTLRLHEVSA 180
 Db 251 QGQAREKHVKEIVISQLTRVHGTSSPYKGLDLSKLHETRLTHRLVSLFNTTLRLHEVSA 310
 Qy 181 QNPNCWMLPLHFRPYISIPVPEQWNNFSTEINTSVLVGPLVSNLEITHTSNLTCKVF 240
 Db 311 QNPNCWMLPLHFRPYISIPVPEQWNNFSTEINTSVLVGPLVSNLEITHTSNLTCKVF 370
 Qy 241 SNTIDTSSQICIRWVTPTRIVCLPSGIFPVCGTSAYHCLNGSSESCFLSFLVPPMTIY 300
 Db 371 SNTITYTNSQICIRWVTPPTQIVCLPSGIFPVCGTSAYHCLNGSSESCFLSFLVPPMTIY 430
 Qy 301 TEQDLYNHVVPKPKHNRKVPILPFVIRAGVLGRGTGIGSTTSTQFYKLSQELNGDMEQ 360
 Db 431 TEQDLYSVISKPRKRVPIPLPFVIRAGVLGRGTGIGSTTSTQFYKLSQELNGDMEQ 490
 Qy 361 VTDSLVTLDQQLNSLAAVLQNRALDILLTAERGCTCLFLGEECCYYVNSGIVTEKVK 420
 Db 491 VADSLVTLDQQLNSLAAVLQNRALDILLTAERGCTCLFLGEECCYYVNSGIVTEKVK 550
 Qy 421 IRDRIQRAEELONTERWGLLSQWMPWLPFLGPLAAILILLFGPCIFNLLVKFVSRI 480
 Db 551 IRDRIQRAEELONTERWGLLSQWMPWLPFLGPLAAILILLFGPCIFNLLVKFVSRI 610
 Qy 481 EAVKLQWLVOME 492
 Db 611 EAVKLQWLVOME 622

RESULT 16

ABP41050

ID ABP41050 standard; protein; 846 AA.

XX


```

Qy 1 MALPYHTFLFTVLLPPALTAPPPCCCTTSSSPYQEFLEXRLPNCNDAPSYRLSKGNS 60
Db 1 MALPYHTFLFTVLLPPALTAPPPCCCTTSSSPYQEFLEXRLPNCNDAPSYRLSKGNS 65
Qy 61 TFTAHTHMPNRCVNSATLCHMANTHYWTGKMINPSCPGGLGATVCWTFYFTHSMDSGGGI 120
Db 61 TFTAHTHMPNRCVNSATLCHMANTHYWTGKMINPSCPGGLGATVCWTFYFTHSMDSGGGI 125
Qy 121 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRLTRVLVSLFNTTLRLHEVSA 180
Db 126 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRLTRVLVSLFNTTLRLHEVSA 185
Qy 181 QNPNCWMLPLHFRPYISIPVPEQWNNFSTEINTTSVLGVLVSNLEIHTHSLTLCVKF 240
Db 186 QNPNCWMLPLHFRPYISIPVPEQWNNFSTEINTTSVLGVLVSNLEIHTHSLTLCVKF 245
Qy 241 SNTIDTSSOCIRWVTPPTTRIVCLPSGIFVCGTSAVHCLNGSSESCFSLFVPPMTIY 300
Db 246 SNTIYTTNSOCIRWVTPPTTRIVCLPSGIFVCGTSAVHCLNGSSESCFSLFVPPMTIY 305
Qy 301 TEQDLYNHVVP-KPHNKRVPILPFPVIRAGVLGRGTGIGTIGTITSTQFYKLSQELNGDME 359
Db 306 TEQDLYSVVISXSPRNKRVPILPFPVIRAGVLGRGTGIGTIGTITSTQFYKLSQELNGDME 365
Qy 360 QVTDLSLVTLODQLNSLAADVLLQNRRLDLLTAKRGGTCLFLGEBRCYVYNQSRIVTEKVK 419
Db 366 QVADSLVTLQDQLNSLAADVLLQNRRLDLLTAKRGGTCLFLGEBRCYVYNQSRIVTEKVK 425
Qy 420 EIRDRIQCRABEELONTERWGLLSQWMPVLPFLGPLAAILLLFGPCIFNLVVKFYSSR 479
Db 426 EIRDRIQCRABEELONTERWGLLSQWMPVLPFLGPLAAILLLFGPCIFNLVVKFYSSR 485
Qy 480 IEAVKQLQWLMQ 492
Db 486 IEAVKQLQWLMQ 498

RESULT 18
ID ADF60170 standard; protein; 544 AA.
AC ADF60170;
XX
XX 12-FEB-2004 (first entry)
XX Human contig polypeptide sequence SEQ ID NO:2537.
XX biological activity; genetic engineering; hybridisation probe; oligomer;
XX primer; chromosome mapping; gene mapping; recombinant protein production;
XX human.
XX Homo sapiens.
XX W02003080795-A2.
XX
XX 02-OCT-2003.
XX
XX 09-AUG-2002; 2002WO-US025485.
XX
XX 09-AUG-2001; 2001US-0311261P.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Yang Y, Wang Z, Weng G, Ma Y;
XX
XX WPI; 2003-876918/81.
XX
XX N-PSDB; ADF59718.
XX
XX New polynucleotides, useful as hybridization probes, oligomers or
XX primers, for chromosome or gene mapping, for the recombinant production
XX of proteins, and for generating antisense DNA or RNA.

```

```

PS Example 3; SEQ ID NO 2537; 571pp; English.
XX
CC The present invention describes isolated polynucleotide sequences (I),
CC which encode polypeptides (II) with biological activity. Also described:
CC (1) a vector comprising (I); (2) an expression vector comprising (I); (3)
CC a host cell genetically engineered to comprise (I) which is operatively
CC associated with a regulatory sequence that modulates expression of (I) in
CC the host cell; (4) a polypeptide (II) encoded by (I); (5) a composition
CC comprising the polypeptide of (4) and a carrier; (6) an antibody directed
CC against the polypeptide of (4); (7) detecting (I) or the polypeptide of
CC (4) in a sample; (8) identifying a compound that binds to the polypeptide
CC of (4); (9) producing the polypeptide of (4); and (10) a collection of
CC polynucleotides comprising at least one of the polynucleotide sequences
CC (I). The polynucleotides (I) can be used as hybridisation probes,
CC oligomers or primers, for chromosome or gene mapping, for the recombinant
CC production of proteins, and for generating antisense DNA or RNA. The
CC present sequence represents a human contig polypeptide sequence, which is
CC used in an example from the present invention.
XX
SQ Sequence 544 AA;

Query Match 87.6%; Score 2308.5; DB 7; Length 544;
Best Local Similarity 88.6%; Pred. No. 1e-199;
Matches 437; Conservative 19; Mismatches 36; Indels 1; Gaps 1;

Qy 1 MALPYHTFLFTVLLPPALTAPPPCCCTTSSSPYQEFLEXRLPNCNDAPSYRLSKGNS 60
Db 6 MALPYHTFLFTVLLPPALTAPPPCCCTTSSSPYQEFLEXRLPNCNDAPSYRLSKGTP 65
Qy 61 TFTAHTHMPNRCVNSATLCHMANTHYWTGKMINPSCPGGLGATVCWTFYFTHSMDSGGGI 120
Db 66 TFTAHTHMPNRCVNSATLCHMANTHYWTGKMINPSCPGGLGATVCWTFYFTHSMDSGGV 125
Qy 121 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRLTRVLVSLFNTTLRLHEVSA 180
Db 126 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRLTRVLVSLFNTTLRLHEVSA 185
Qy 181 QNPNCWMLPLHFRPYISIPVPEQWNNFSTEINTTSVLGVLVSNLEIHTHSLTLCVKF 240
Db 186 QNPNCWMLPLHFRPYISIPVPEQWNNFSTEINTTSVLGVLVSNLEIHTHSLTLCVKF 245
Qy 241 SNTIDTSSOCIRWVTPPTTRIVCLPSGIFVCGTSAVHCLNGSSESCFSLFVPPMTIY 300
Db 246 SNTIYTTNSOCIRWVTPPTTRIVCLPSGIFVCGTSAVHCLNGSSESCFSLFVPPMTIY 305
Qy 301 TEQDLYNHVVP-KPHNKRVPILPFPVIRAGVLGRGTGIGTIGTITSTQFYKLSQELNGDME 359
Db 306 TEQDLYSVVISXSPRNKRVPILPFPVIRAGVLGRGTGIGTIGTITSTQFYKLSQELNGDME 365
Qy 360 QVTDLSLVTLODQLNSLAADVLLQNRRLDLLTAKRGGTCLFLGEBRCYVYNQSRIVTEKVK 419
Db 366 QVADSLVTLQDQLNSLAADVLLQNRRLDLLTAKRGGTCLFLGEBRCYVYNQSRIVTEKVK 425
Qy 420 EIRDRIQCRABEELONTERWGLLSQWMPVLPFLGPLAAILLLFGPCIFNLVVKFYSSR 479
Db 426 EIRDRIQCRABEELONTERWGLLSQWMPVLPFLGPLAAILLLFGPCIFNLVVKFYSSR 485
Qy 480 IEAVKQLQWLMQ 492
Db 486 IEAVKQLQWLMQ 498

RESULT 19
AAW99345
ID AAW99345 standard; protein; 538 AA.
XX
XX AAW99345;
AC AAW99345;
XX
XX 21-MAY-1999 (first entry)
XX
XX Human endogenous retrovirus W ORF1.
XX
XX Clone; human endogenous retrovirus; genome; autoimmune disease;

```

KW multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KW disseminated lupus erythematosus; pregnancy; chromosomal marker.
XX
OS Human endogenous retrovirus.
XX
XX WO9902696-A1.
XX
XX 21-JAN-1999.
XX
XX 06-JUL-1998; 98WO-FR001442.
XX
XX 07-JUL-1997; 97FR-00008815.
XX
XX (INMR) BIO MERIEUX.
XX
XX Beseme F, Blond J, Bouton O, Mandrand B, Mallet F;
PI
XX WPI; 1999-120897/10.
XX
XX New nucleic acid sequences from human endogenous retrovirus-W - expressed
PT exclusively in placenta and useful in diagnosis and therapy of autoimmune
PT disease, and abnormal or failed pregnancy.
XX
XX Example 8; Fig 6; 106pp; French.
XX
XX This sequence represents the open reading frame 1 encoding the env
CC protein from the human endogenous retrovirus (HERV) W (AA25665). Nucleic
CC acids, their fragments or peptides encoded by them derived from the HERV-
CC W genome are markers of autoimmune disease (e.g. multiple sclerosis,
CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
CC dependent diabetes and related pathologies) and of abnormal or
CC unsuccessful pregnancy and can be used as chromosomal markers for
CC susceptibility to these conditions, or proximity markers of genes
CC associated with this susceptibility
XX
SQ Sequence 538 AA;

Query Match 84.9%; Score 2237; DB 2; Length 538;
Best Local Similarity 85.6%; Pred. No. 3e-193;
Matches 421; Conservative 26; Mismatches 45; Indels 0; Gaps 0;
Qy 1 MALPYHTFLTVLPPALFAPPCCCTTSSPQVEFLXTRLPDGNIDAPSYRSLSGNS 60
Db 1 MGLPYHIFLCSVSPCTLTAPPCCRCWTSSPHPEFLWRMQRPDGNIDAPSYRSLSGTP 60
Qy 61 TFTAHTMPNCYNSATLQWANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMDSGGI 120
Db 61 TFTAHTMPNCYNSATLQWANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMDSGGV 120
Qy 121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLEHETLRLVSLFNTTLRLHEVSA 180
Db 121 QGQAREKHXVEVISQLTGVHGTSSPYKGLDLKLEHETLRLVSLFNTTLRLHEVSA 180
Qy 181 QNPNCWMLPLHFRPYISIPVPOWNNFSTEINTSVLGPLVSNLEITHTSNLTCKVF 240
Db 181 QNPNCWICLPLNFRPYISIPVPOWNNFSTEINTSVLGPLVSNLEITHTSNLTCKVF 240
Qy 241 SNTIDTSSQCIWRWTPTRIVCLPSGIFFCVCGTSAYHCLNGSSSESCFLSFLVPPMTIY 300
Db 241 SNTITYTNSQCIWRWTPTRIVCLPSGIFFCVCGTSAYHCLNGSSSESCFLSFLVPPMTIY 300
Qy 301 TEQDLYNHVVPKPKKRVPIPLPFVIRAGVLRGLTGIGSTTSTQFYKLSQELNGDMEQ 360
Db 301 TEQDLYSVITSKPKKRVPIPLPFVIRAGVLRGLTGIGSTTSTQFYKLSQELNGDMEQ 360
Qy 361 VTDLSVLTQDLNLSAAVQLNRRALDILLAKRGCTCLFLGEEECYVYVNSRIIVTEKVE 420
Db 361 VADLSVLTQDLNLSAAVQLNRRALDILLAKRGCTCLFLGEEECYVYVNSGIIVTEKVE 420
Qy 421 IRDRIQCRAELONTERWGLLSQWMPVLPGLPLAAILLLLPGLPCITENLLVVFVSSRI 480
Db 421 IPDRIQRIABELRNTGWLSSRWMPVLPGLPLAAILLLLPGLPCITENLLVVFVSSRI 480

Qy 481 EAVKLOMVLQME 492
Db 481 EAVKLOMPEKMQ 492

RESULT 20
AAM18008
ID AAM18008 standard; protein; 263 AA.
XX
XX AAM18008;
AC
XX 12-OCT-2001 (first entry)
DT
XX
DE
XX Peptide #4442 encoded by probe for measuring cervical gene expression.
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX
XX Homo sapiens.
OS
XX WO200157278-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US000670.
PF
XX
XX 04-FEB-2000; 2000US-0180312P.
PR
XX 26-MAY-2000; 2000US-0207456P.
PR
XX 30-JUN-2000; 2000US-00608408.
PR
XX 03-AUG-2000; 2000US-00632366.
PR
XX 21-SEP-2000; 2000US-0234687P.
PR
XX 27-SEP-2000; 2000US-0236359P.
PR
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA

XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human cervical epithelial cells.
XX
XX Claim 27; SEQ ID NO 22834; 487pp; English.
XX

CC The present invention relates to human single exon nucleic acid probes
CC (SENPs: see AAI10068-AAI28459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human Hela cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 263 AA;

Query Match 46.9%; Score 1235.5; DB 4; Length 263;
Best Local Similarity 89.4%; Pred. No. 5.1e-103;
Matches 236; Conservative 8; Mismatches 19; Indels 1; Gaps 1;

Qy 126 EKQVKEAISQLTRGHSTPSPYKGLVLSKLEHETLRLVSLFNTTLRLHEVSAQNP 185
Db 1 EKHKVEISQLTWVHSTPGYKGLDLKLEHETLRLVSLFNTTLRLHEVSAQNP 60
Qy 186 CWMLPLHFRPYISIPVPOWNNFSTEINTSVLGPLVSNLEITHTSNLTCKVFSNTID 245
Db 61 CWMLPLHFRPYISIPVPOWNNFSTEINTSVLGPLVSNLEITHTSNLTCKVFSNTID 119
Qy 246 TTSOICRWTPTRIVCLPSGIFFCVCGTSAYHCLNGSSSESCFLSFLVPPMTIYTBODL 305
Db 120 TANSQCIRWTPTRIVCLPSGIFFCVCGTSAYHCLNGSSSESCFLSFLVPPMTIYTBODL 179

```

Qy 306 YNHVVPKPHNKRVPILPFVIRAGVLGRGTGIGTITSTQFYKLSQEIINGDMQVTDLSL 365
Db 180 YNHVIPKPRNKRVPILPFVIRAGVLGRGTGIGTITSTQFYKLSQELANGDMQVADTDL 239
Qy 366 VTLDQQLNSLAADVQLNRRALDILL 389
Db 240 VTLDQQLNSLAADVQLNRRALDILL 263

RESULT 21
ID ABB37043 standard; peptide; 263 AA.
AC ABB37043;
XX
XX 04-FEB-2002 (first entry)
XX
XX Peptide #4549 encoded by human foetal liver single exon probe.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
XX Homo sapiens.
XX
XX WO200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000669.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX
XX 26-MAY-2000; 2000US-0207456P.
XX
XX 30-JUN-2000; 2000US-00608408.
XX
XX 03-AUG-2000; 2000US-00632366.
XX
XX 21-SEP-2000; 2000US-0234687P.
XX
XX 27-SEP-2000; 2000US-0236359P.
XX
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human fetal liver.
XX
XX Claim 27; SEQ ID NO 29678; 639pp + Sequence Listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for measuring
XX human gene expression in a sample derived from human foetal liver. The
XX single exon nucleic acid probes may be used for predicting, measuring and
XX displaying gene expression in samples derived from human fetal liver. The
XX present sequence is a peptide encoded by a single exon nucleic acid probe
XX of the invention. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 263 AA;
XX
XX Query Match 46.9%; Score 1235.5; DB 4; Length 263;
XX Best Local Similarity 89.4%; Pred. No. 5.1e-103;
XX Matches 236; Conservative 8; Mismatches 19; Indels 1; Gaps 1;
XX
Qy 126 EKOVKEAISOLTRGHSTPSPYKGLVLSKLHETLTHTRLVSLFNTTLTRLHVEVSAQNPTN 185
Db 1 EKHIKEVISQTLVWHSTPGPYKGLDLSKLHETLTHTRLVSLFNTTLTGLHVEVSAQNPTN 60
Qy 186 CWMCLPLHFRPYISIPVPEQNNFSTEINTTSLVGLPLVSNLEITHTSNLTCVKFSNTID 245
Db 61 CWMCLPLHFRPYISIPVPEQNNFSTEINTTSLVGLPL-SNLEITHTSNLTCVKFSNTID 119
Qy 246 TTSSQCIRWTPTRIVCLPSGIFVCGTSAYHCLNGSSSVCFLSFLVAPMPYITTDQL 305

```

```

Db 120 TANSQCIRWTPTRIVCLPSGIFVCGTSAYHCLNGSSSVCFLSFLVAPMPYITTDQL 179
Qy 306 YNHVVPKPHNKRVPILPFVIRAGVLGRGTGIGTITSTQFYKLSQEIINGDMQVTDLSL 365
Db 180 YNHVIPKPRNKRVPILPFVIRAGVLGRGTGIGTITSTQFYKLSQELANGDMQVADTDL 239
Qy 366 VTLDQQLNSLAADVQLNRRALDILL 389
Db 240 VTLDQQLNSLAADVQLNRRALDILL 263

RESULT 22
ID AAM30519 standard; protein; 263 AA.
XX
XX AAM30519;
XX
XX 17-OCT-2001 (first entry)
XX
XX Peptide #4556 encoded by probe for measuring placental gene expression.
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000663.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX
XX 26-MAY-2000; 2000US-0207456P.
XX
XX 30-JUN-2000; 2000US-00608408.
XX
XX 03-AUG-2000; 2000US-00632366.
XX
XX 21-SEP-2000; 2000US-0234687P.
XX
XX 27-SEP-2000; 2000US-0236359P.
XX
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX
XX Claim 27; SEQ ID NO 30788; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP:
XX see AAI31315-AAI57546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders
XX
XX Sequence 263 AA;
XX
XX Query Match 46.9%; Score 1235.5; DB 4; Length 263;
XX Best Local Similarity 89.4%; Pred. No. 5.1e-103;
XX Matches 236; Conservative 8; Mismatches 19; Indels 1; Gaps 1;
XX
Qy 126 EKOVKEAISOLTRGHSTPSPYKGLVLSKLHETLTHTRLVSLFNTTLTRLHVEVSAQNPTN 185
Db 1 EKHIKEVISQTLVWHSTPGPYKGLDLSKLHETLTHTRLVSLFNTTLTGLHVEVSAQNPTN 60
Qy 186 CWMCLPLHFRPYISIPVPEQNNFSTEINTTSLVGLPLVSNLEITHTSNLTCVKFSNTID 245
Db 61 CWMCLPLHFRPYISIPVPEQNNFSTEINTTSLVGLPL-SNLEITHTSNLTCVKFSNTID 119

```


CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 263 AA;

Query Match 46.9%; Score 1235.5; DB 4; Length 263;
Best Local Similarity 89.4%; Pred. No. 5.1e-103;
Matches 236; Conservative 8; Mismatches 19; Indels 1; Gaps 1;
QY 126 EKQVKEAISQTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTN 185
Db 1 EKHKEIVISQLTWVHSTPGPYKGLDLSKLHETLHTHTGLVSLFNTTLTGLHEVSAQNPTN 60
QY 186 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLGVLVSNLEITHTSNTLTCVKFSNTID 245
Db 61 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLGVLVSNLEITHTSNTLTCVKFSNTID 119
QY 246 TTSQCIKRWTPTRIVCLPSGIFVCGTSAYHCLNGSSSVMCFSLFVLPMTIYTEQDL 305
Db 120 TANSQCIKRWTPTRIVCLPSGIFVCGTSAYHCLNGSSSVMCFSLFVLPMTIYTEQDL 179
QY 306 YNHVVPKPNKRPVLPFVIRAGVLGRGTGIGSITTSQFYKLSQELNGDMQVTDLS 365
Db 180 YNHVVPKPNKRPVLPFVIRAGVLGRGTGIGSITTSQFYKLSQELNGDMQVTDLS 239
QY 366 VTLDQQLNSLAADVQLNRRALDIL 389
Db 240 VTLDQQLNSLAADVQLNRRALDIL 263

RESULT 25
AAW70185
ID AAW70185 standard; protein; 263 AA.
XX
AC AAW70185;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 30491.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US0000668.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX

PS Example 4; SEQ ID NO 30491; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention
XX
SQ Sequence 263 AA;

Query Match 46.9%; Score 1235.5; DB 4; Length 263;
Best Local Similarity 89.4%; Pred. No. 5.1e-103;
Matches 236; Conservative 8; Mismatches 19; Indels 1; Gaps 1;
QY 126 EKQVKEAISQTRGHSTPSPYKGLVLSKLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTN 185
Db 1 EKHKEIVISQLTWVHSTPGPYKGLDLSKLHETLHTHTGLVSLFNTTLTGLHEVSAQNPTN 60
QY 186 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLGVLVSNLEITHTSNTLTCVKFSNTID 245
Db 61 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLGVLVSNLEITHTSNTLTCVKFSNTID 119
QY 246 TTSQCIKRWTPTRIVCLPSGIFVCGTSAYHCLNGSSSVMCFSLFVLPMTIYTEQDL 305
Db 120 TANSQCIKRWTPTRIVCLPSGIFVCGTSAYHCLNGSSSVMCFSLFVLPMTIYTEQDL 179
QY 306 YNHVVPKPNKRPVLPFVIRAGVLGRGTGIGSITTSQFYKLSQELNGDMQVTDLS 365
Db 180 YNHVVPKPNKRPVLPFVIRAGVLGRGTGIGSITTSQFYKLSQELNGDMQVTDLS 239
QY 366 VTLDQQLNSLAADVQLNRRALDIL 389
Db 240 VTLDQQLNSLAADVQLNRRALDIL 263

RESULT 26
AAW57770
ID AAW57770 standard; protein; 263 AA.
XX
AC AAW57770;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 29875.
XX
KW Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US0000667.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains.
XX

```
XX CC The present invention provides a number of single exon nucleic acid
PS CC probes which are derived from genomic sequences expressed in the human
XX CC brain. They can be used to measure gene expression in brain cell samples,
CC CC which may enable the diagnosis and improved treatment of nervous system
CC CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC CC the probes of the invention
XX CC
XX CC Sequence 263 AA;
SQ Query Match 46.9%; Score 1235.5; DB 4; Length 263;
Best Local Similarity 89.4%; Pred. No. 5.1e-103;
Matches 236; Conservative 8; Mismatches 19; Indels 1; Gaps 1;
QY 126 EKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLTHTRIVLSLNTLTGLHEVSAQNPTN 185
DB 1 EKHKEVISQLTWVHSTPGPYKGLDLSKLHETLTHTRIVLSLNTLTGLHEVSAQNPTN 60
QY 186 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGLVSNLEITHTSNLTVCVKSNTID 245
DB 61 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGLVSNLEITHTSNLTVCVKSNTID 119
QY 246 TTSSQCIRWVTPPTTRIVCLPSGIFVCGTSAYHCLNGSSCMFLSLVPPMTIYTEQDL 305
DB 120 TANSQCIRWVTPPTTRIVCLPSGIFVCGTSAYHCLNGSSCMFLSLVPPMTIYTEQDL 179
QY 306 YNHVVPKPNKRVPIILPFVIRAGVLRGLGTGIGSIITSTQFYKLSQELNGDMQVTDLSL 365
DB 180 YNHVIPKRNKRVPIILPFVIRAGVLRGLGTGIGSIITSTQFYKLSQELNGDMQVTDLSL 239
QY 366 VTLODQLNSLAADVQLNRRALDLL 389
DB 240 VTLODQLNSLAADVQLNRRALDLL 263
RESULT 27
ABG51885
ID ABG51885 standard; peptide; 263 AA.
XX AC ABG51885;
XX DT 25-FEB-2003 (first entry)
XX DE Human liver peptide, SEQ ID NO 30533.
XX KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX KW hypercholesterolaemia; coronary heart disease.
XX OS Homo sapiens.
XX EN WO200157273-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000664.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488898/53.
XX XX
```

```
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
XX Claim 27; SEQ ID NO 30533; 658pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABG47348-ABG5930 represent human
CC liver single exon encoded peptides of the invention. Note: The sequence
CC information for this patent does not appear in the printed specification
CC but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 263 AA;
SQ Query Match 46.9%; Score 1235.5; DB 4; Length 263;
Best Local Similarity 89.4%; Pred. No. 5.1e-103;
Matches 236; Conservative 8; Mismatches 19; Indels 1; Gaps 1;
QY 126 EKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLTHTRIVLSLNTLTGLHEVSAQNPTN 185
DB 1 EKHKEVISQLTWVHSTPGPYKGLDLSKLHETLTHTRIVLSLNTLTGLHEVSAQNPTN 60
QY 186 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGLVSNLEITHTSNLTVCVKSNTID 245
DB 61 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLVGLVSNLEITHTSNLTVCVKSNTID 119
QY 246 TTSSQCIRWVTPPTTRIVCLPSGIFVCGTSAYHCLNGSSCMFLSLVPPMTIYTEQDL 305
DB 120 TANSQCIRWVTPPTTRIVCLPSGIFVCGTSAYHCLNGSSCMFLSLVPPMTIYTEQDL 179
QY 306 YNHVVPKPNKRVPIILPFVIRAGVLRGLGTGIGSIITSTQFYKLSQELNGDMQVTDLSL 365
DB 180 YNHVIPKRNKRVPIILPFVIRAGVLRGLGTGIGSIITSTQFYKLSQELNGDMQVTDLSL 239
QY 366 VTLODQLNSLAADVQLNRRALDLL 389
DB 240 VTLODQLNSLAADVQLNRRALDLL 263
RESULT 28
NAM05648
ID NAM05648 standard; protein; 263 AA.
XX AC NAM05648;
XX DT 09-OCT-2001 (first entry)
XX DE Peptide #4330 encoded by probe for measuring breast gene expression.
XX KW Probe; human; breast disease; breast cancer; development disorder;
XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX OS Homo sapiens.
XX XX WO200157270-A2.
XX PD 09-AUG-2001.
XX PF 29-JAN-2001; 2001WO-US000661.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
```

```
PR 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX Novel single exon nucleic acid probe used to measuring gene expression in
XX a human breast.
XX Claim 27; SEQ ID NO 14388; 322pp; English.
XX The present invention relates to novel single exon nucleic acid probes
XX (see AA100010-AA110067). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for measuring human gene expression in
XX a human breast sample, where the probe hybridizes at high stringency to a
XX nucleic acid expressed in the human breast. The probes are useful for
XX predicting, diagnosing, grading, staging, monitoring and prognosing
XX diseases of the human breast, particularly those diseases with polygenic
XX aetiology. The diseases include: breast cancer, disorders of development,
XX inflammatory diseases of the breast, fibrocystic changes, proliferative
XX breast disease and non-carcinoma tumours. Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 263 AA;

Query Match 46.9%; Score 1235.5; DB 4; Length 263;
Best Local Similarity 89.4%; Pred. No. 5.1e-103;
Matches 236; Conservative 8; Mismatches 19; Indels 1; Gaps 1;

QY 126 EKQVKEAISQLTGRCHSTSPYKGLVLSKLTHTLTHRLVSLFNTTLRLHEVSAQNPTN 185
DB 1 EKHIKEVISQLTWHVSTPGPYKGLDLSKLTHTLTHRLVSLFNTTLRLHEVSAQNPTN 60
QY 186 CWMCLPLHFRPYISIPVPEQWNNFSTINTSVLGLVNSLNTLTSTNLTCTVKFSNTID 245
DB 61 CWMCLPLHFRPYISIPVPEQWNNFSTINTSVLGLVNSLNTLTSTNLTCTVKFSNTID 119
QY 246 TTSSQCIIRWTPPRIVCLPSGIFPVCGTSAYHCLNGSSSMCPFLSPVPPMTIYTBQDL 305
DB 120 TANSQCIIRWTPPRIVCLPSGIFPVCGTSAYHCLNGSSSMCPFLSPVPPMTIYTBQDL 179
QY 306 YNHVVPKPNKRVPIPLFVIRAGVLRLTGIGTSITSTQFYKLSQELNGMEQVTDLSL 365
DB 180 YNHVVPKPNKRVPIPLFVIRAGVLRLTGIGTSITSTQFYKLSQELNGMEQVTDLSL 239
QY 366 VTLDQQLNSLAADVQLNRRALDLL 389
DB 240 VTLDQQLNSLAADVQLNRRALDLL 263

RESULT 29
ABG39819
ID ABG39819 standard; peptide; 263 AA.
XX AC ABG39819;
XX AC ABG39819;
DT 19-AUG-2002 (first entry)
DE Human peptide encoded by genome-derived single exon probe SEQ ID 29484.
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenar syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
```

hyaline membrane disease.

Homo sapiens.

WO200186003-A2.

15-NOV-2001.

30-JAN-2001; 2001WO-US000665.

04-FEB-2000; 2000US-0180312P.

26-MAY-2000; 2000US-0207456P.

30-JUN-2000; 2000US-00608408.

03-AUG-2000; 2000US-00632366.

21-SEP-2000; 2000US-0234687P.

27-SEP-2000; 2000US-0236359P.

04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2002-114183/15.

Spatially-addressable set of single exon nucleic acid probes, used to

measure gene expression in human lung samples.

Claim 27; SEQ ID NO 29484; 634pp; English.

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridise at high stringency to a nucleic acid expressed in the human lung; measuring gene expression in a sample derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array, identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karagenar syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention. Note: the sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 263 AA;

Query Match 46.9%; Score 1235.5; DB 5; Length 263;

Best Local Similarity 89.4%; Pred. No. 5.1e-103;

Matches 236; Conservative 8; Mismatches 19; Indels 1; Gaps 1;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2005, 21:09:06 ; Search time 45 Seconds
(without alignments)
726.551 Million cell updates/sec

Title: US-09-319-156B-10
Perfect score: 2634
Sequence: 1 MALPYTFTFLTVLLPPFALT.....KFVSSRIEAVKQVLQMP 493

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PGTUS COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2632	99.9	493	4	US-08-979-847B-106
2	2319	88.0	538	3	US-09-175-528-4
3	559.5	21.2	567	1	US-08-007-282B-2
4	516	19.6	162	4	US-08-979-847B-109
5	452	17.2	433	3	US-08-691-563C-87
6	452	17.2	433	4	US-09-374-766-87
7	452	17.2	433	4	US-08-979-847B-81
8	407.5	15.5	632	4	US-09-315-127-2
9	407.5	15.5	632	4	US-09-315-127-3
10	407.5	15.5	665	4	US-09-309-572-14
11	407.5	15.5	665	4	US-09-718-096-14
12	407.5	15.5	1312	4	US-09-554-572-26
13	389.5	14.8	673	3	US-09-075-272-5
14	385	14.6	654	4	US-09-315-127-11
15	385	14.6	654	4	US-09-315-127-12
16	376.5	14.3	645	4	US-09-315-127-8
17	376.5	14.3	645	4	US-09-315-127-9
18	376.5	14.3	660	3	US-09-111-085-2
19	376.5	14.3	660	3	US-09-376-781-5
20	368	14.0	638	3	US-09-376-781-6
21	364	13.8	192	3	US-08-486-099-117
22	364	13.8	192	3	US-08-360-107A-117
23	364	13.8	192	3	US-08-484-223B-107
24	364	13.8	192	3	US-08-919-597-107
25	364	13.8	192	3	US-08-475-668A-107
26	364	13.8	192	3	US-08-485-551A-107
27	364	13.8	192	3	US-08-471-913A-107

28	364	13.8	192	3	US-08-485-264A-107	Sequence 107, App
29	364	13.8	192	3	US-08-474-349A-107	Sequence 107, App
30	364	13.8	192	4	US-08-470-896-107	Sequence 107, App
31	364	13.8	192	4	US-08-485-546A-107	Sequence 5, Appli
32	357.5	13.6	667	4	US-09-315-127-5	Sequence 6, Appli
33	357.5	13.6	667	4	US-09-315-127-6	Sequence 13, Appli
34	357.5	13.6	667	4	US-09-070-630-13	Sequence 10, Appli
35	352	13.4	232	4	US-09-555-352-10	Sequence 4, Appli
36	349	13.2	657	3	US-09-111-085-4	Sequence 4, Appli
37	349	13.2	657	3	US-09-376-781-4	Sequence 10, Appli
38	347	13.2	657	4	US-09-171-553B-10	Sequence 3, Appli
39	345	13.1	656	3	US-09-376-781-3	Sequence 6, Appli
40	345	13.1	656	4	US-09-171-553B-6	Sequence 1593, Ap
41	341	12.9	191	4	US-09-350-841A-1593	Sequence 7, Appli
42	311	11.8	196	2	US-08-484-126-7	Sequence 49, Appli
43	311	11.8	196	4	US-09-374-909-7	Sequence 3, Appli
44	298	11.3	67	1	US-08-686-878A-49	
45	298	11.3	67	1	US-08-721-489-3	

ALIGNMENTS

RESULT 1
US-08-979-847B-106
; Sequence 106, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; BESEME, FREDERIC
; BEDIN, FREDERIC
; PARANHOS-BACCALA, GLAUCIA
; KOMURIAN-PRADEL, FLORENCE
; JOLIVET-REYNAUD, COLETTE
; MANDRAND, BERNARD
; GARSON JEREMY
; TUNE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
; THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,847B
; FILING DATE: 26-No. 6582703-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BERRIDGE, WILLIAM P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 39046A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 493 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 106:
US-08-979-847B-106

Query Match 99.9%; Score 2632; DB 4; Length 493;
 Best Local Similarity 100.0%; Pred. No. 2.8e-262;
 Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFELXRLTRPGNIDAPSYSLSGNS 60
 DB 1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFELXRLTRPGNIDAPSYSLSGNS 60
 QY 61 TFTAHTMPNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMDGGI 120
 DB 61 TFTAHTMPNCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMDGGI 120
 QY 121 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRLTHRLVSLFNTTLRLHEVSA 180
 DB 121 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRLTHRLVSLFNTTLRLHEVSA 180
 QY 181 QNPTNCMCLPLHFRPYISIPVPEQNNFSTEINTTSVLGVLVSNLEITHSNLTCVKF 240
 DB 181 QNPTNCMCLPLHFRPYISIPVPEQNNFSTEINTTSVLGVLVSNLEITHSNLTCVKF 240
 QY 241 SNTIDTSSOCIRWVTPTRIVCLPSGIFVCGTSAVHCLNGSSESCFSLFVPPMTIY 300
 DB 241 SNTIDTSSOCIRWVTPTRIVCLPSGIFVCGTSAVHCLNGSSESCFSLFVPPMTIY 300
 QY 301 TEQDLYNHVVPKPHNRKVPILPFVIRAGVLRGLTGIGSITTSQFYKLSQEBINGDMEQ 360
 DB 301 TEQDLYNHVVPKPHNRKVPILPFVIRAGVLRGLTGIGSITTSQFYKLSQEBINGDMEQ 360
 QY 361 VTDLSVLTQDLSLAAVLQNRALDILLTAKRGCTCLFGEERCYYVNSRIVTEKVK 420
 DB 361 VTDLSVLTQDLSLAAVLQNRALDILLTAKRGCTCLFGEERCYYVNSRIVTEKVK 420
 QY 421 IRDRIQRAEQLONTERWGLLSQMPWLPFLGFLAALILLLLFGPCIFNLLVKFVSSRI 480
 DB 421 IRDRIQRAEQLONTERWGLLSQMPWLPFLGFLAALILLLLFGPCIFNLLVKFVSSRI 480
 QY 481 EAVKLQWVLMQE 492
 DB 481 EAVKLQWVLMQE 492

RESULT 3

US-08-007-282B-2
 ; Sequence 2, Application US/08007282B
 ; Patent No. 5403582
 ; GENERAL INFORMATION:
 ; APPLICANT: NAZERIAN, KEYVAN
 ; APPLICANT: CALVERT, JAY G.
 ; APPLICANT: WITTER, RICHARD L.
 ; APPLICANT: YAMAGIDA, NOBORU
 ; TITLE OF INVENTION: VACCINE COMPRISING FOWLPOX VIRUS
 ; TITLE OF INVENTION: RECOMBINANTS EXPRESSING THE ENVELOPE GLYCOPROTEIN OF AN
 ; TITLE OF INVENTION: AVIAN RETICULOENDOTHELIOSIS RETROVIRUS
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
 ; STREET: 8110 Gatehouse Road Suite 500 East
 ; CITY: Falls Church
 ; STATE: VA
 ; COUNTRY: USA
 ; ZIP: 22042
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/007,282B
 ; FILING DATE: 19930121
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: MURPHY JR., GERALD M.
 ; REGISTRATION NUMBER: 28,977
 ; REFERENCE/DOCKET NUMBER: 1644-104P
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 703-205-8000
 ; TELEFAX: 703-205-8050
 ; TELEX: 248345

RESULT 2

US-09-175-928-4
 ; Sequence 4, Application US/09175928A
 ; Patent No. 6312921
 ; GENERAL INFORMATION:
 ; APPLICANT: Jacobs, Kenneth
 ; APPLICANT: McCoy, John M.
 ; APPLICANT: LaVallie, Edward R.
 ; APPLICANT: Collins-Racie, Lisa A.
 ; APPLICANT: Evans, Cheryl
 ; APPLICANT: Merberg, David
 ; APPLICANT: Treacy, Maurice
 ; APPLICANT: Mi, Sha
 ; APPLICANT: Genetics Institute, Inc.
 ; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
 ; FILE REFERENCE: 6006B AJ172A
 ; CURRENT APPLICATION NUMBER: US/09/175,928A
 ; CURRENT FILING DATE: 1998-10-20
 ; NUMBER OF SEQ ID NOS: 62
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 4
 ; LENGTH: 538
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-175-928-4

Query Match 88.0%; Score 2319; DB 3; Length 538;
 Best Local Similarity 88.6%; Pred. No. 5.9e-230;
 Matches 436; Conservative 18; Mismatches 38; Indels 0; Gaps 0;

QY 1 MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFELXRLTRPGNIDAPSYSLSGNS 60

```
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 567 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-007-282B-2

Query Match 21.2%; Score 559.5; DB 1; Length 567;
Best Local Similarity 31.9%; Pred. No. 1.5e-48;
Matches 153; Conservative 75; Mismatches 158; Indels 93; Gaps 19;

QY 66 THPRNCYSATLCMHANTHYWT-----GKMINPSCPGGLGATVCWTFY 110
DB 121 TOMHSTCYKTEQCTLLGKTYFTAILQKTLGSDYEDGNKLIQASCTGTGKPVCDPVA 180
QY 111 HTSMDSGGIGOGAREKQKVEAISQLTRGHSTPS-PYKGLVLSKLHETLTHRLVSLFN 169
DB 181 PTVVSDGGGPTDMIREESVRELEEIIIR-HSYPSVQVHPLALPR-----SRGVDLDP 231
QY 170 TT---LTRLHEV-SAONP---TNCMCLPLHFRPYISIPVPEQWN-----NFGTEI--- 213
DB 232 QTSIDLEATHQVLNATNPKLAENCLWLTLL-GTPIPAAIPTNGNVTLDCNCSLSLPFGC 289
QY 214 -----NTTSVLVGLVSNLEITHTSNLTCTVKFSNTIDTTSQCIRWVTP 257
DB 290 NPPGSDIVSCVAGEADNRTGIPVG-----YVHFTNCTSIQ---EVTNETSQ---MGN 335
QY 258 PTRIVCLPSGIFVCGTS-AYHCLNGSSESMCFSLFVPPMTYITEQDLNHNHVPKP--- 313
DB 336 LTRL-CPPPGHFVFCGNNMAYTALPNKWIGLICILASIVPDISISGEE---PIPLFSIE 390
QY 314 -----HNKRYPILPFVIRAGVLGRL---GTGIGSITTSOTFYKLSOEINGMEQVDSL 365
DB 391 YTARRHKRAVQFPLLVGLGSGATLAGGTG---VSVITYHKLSQLIEDVQALSGTI 447
QY 366 VTLDQQLNSLAAYVLQNRRLDLLTAKRGCTCLFLGEERCYVYNQSRIVTEKVKIEDRI 425
DB 448 NDLQDQIDSLAEVVLQNRRLDLLTAEQGGICLALQKCCFYANKSGIVDRKIRKQEDL 507
QY 426 QCRAEELQNTERGKLSQWMPWVLPFLGPLAAIILLFPGCIFNLLVKFVSSRIEAVK 484
DB 508 IERKRALYDNLMSGLNGFLPYLLPLGLPLGLILFTLGLPCIMKTLTRIHKIQAVK 566

RESULT 4
US-08-979-847B-109
; Sequence 109, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; BESEME, FREDERIC
; BEDIN, FREDERIC
; PARANHOS-BACCALA, GLAUCIA
; KOMURIAN-PRADEL, FLORENCE
; JOLIVET-REYNAUD, COLETTE
; MANDRAND, BERNARD
; GARSON, JEREMY
; TUKE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
; THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,563C
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,847B
; FILING DATE: 26-No. 6582703-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BERRIDGE, WILLIAM P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 39046A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 162 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 109:
US-08-979-847B-109

Query Match 19.6%; Score 516; DB 4; Length 162;
Best Local Similarity 87.6%; Pred. No. 6.4e-45;
Matches 99; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 381 QNRRLDLLTAKRGCTCLFLGEERCYVYNQSRIVTEKVKIEDRIQCRAEELQNTERGWL 440
DB 1 QNRRLDLLTAKRGCTCLFLGEERCYVYNQSGIITKVKIEIXDRIKRAEDLQNTAPWGL 60
QY 441 LSOQWMPWVLPFLGPLAAIILLFPGCIFNLLVKFVSSRIEAVKQLQVLQMEP 493
DB 61 LSOQWMPWVLPFLGPLAAIILLFPGCIFNLLVKFVSSRIEAVKQLQVLQMEP 113

RESULT 5
US-08-691-563C-87
; Sequence 87, Application US/08691563C
; Patent No. 6001987
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glauca PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
; THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,563C
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
```

```
;
; REFERENCE/DOCKET NUMBER: WPB 38588
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-691-563C-87

Query Match          17.2%; Score 452; DB 3; Length 433;
Best Local Similarity 54.3%; Pred. No. 1.2e-37;
Matches 89; Conservative 27; Mismatches 38; Indels 10; Gaps 2;

QY 334 GTGIGSTTTQFYKLSQINGMEQVTDLSVTLQDQLNSLAAVLVQNRRLDILLTAKR 393
Db 6 GTGIAGLSTLSYHTLSKNFSDLSQEIIMKSIITLQSDLSLAAMTLQNRGPHLLTAEK 65

QY 394 GTGCTFLGECRCYVYNQSRIVTEKVKIIRDRI---QCRAEELQNTERWGLLSQMPWVL 449
Db 66 GGLCTFLGECRCYVYNQSGIVRDATWHLQERASDIRQCLNSYTN-----LMSWATWLL 119

QY 450 PFLGPLAAILLLLFPCIFNLVVKFVSSRIEAVKQLQWLQMEP 493
Db 120 PFLGPMMAIILLTTFGCFIKLLVVKFVSSRIEAIKQLQWLQMEP 163

RESULT 6
US-09-374-766-87
; Sequence 87, Application US/09374766
; Patent No. 6579526
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glaucia PARANHOS-BACCALA
; APPLICANT: Glaucia KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/374,766
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,563
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38588
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
```

```
;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-374-766-87

Query Match          17.2%; Score 452; DB 4; Length 433;
Best Local Similarity 54.3%; Pred. No. 1.2e-37;
Matches 89; Conservative 27; Mismatches 38; Indels 10; Gaps 2;

QY 334 GTGIGSTTTQFYKLSQINGMEQVTDLSVTLQDQLNSLAAVLVQNRRLDILLTAKR 393
Db 6 GTGIAGLSTLSYHTLSKNFSDLSQEIIMKSIITLQSDLSLAAMTLQNRGPHLLTAEK 65

QY 394 GTGCTFLGECRCYVYNQSRIVTEKVKIIRDRI---QCRAEELQNTERWGLLSQMPWVL 449
Db 66 GGLCTFLGECRCYVYNQSGIVRDATWHLQERASDIRQCLNSYTN-----LMSWATWLL 119

QY 450 PFLGPLAAILLLLFPCIFNLVVKFVSSRIEAVKQLQWLQMEP 493
Db 120 PFLGPMMAIILLTTFGCFIKLLVVKFVSSRIEAIKQLQWLQMEP 163

RESULT 7
US-08-979-847B-81
; Sequence 81, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
; APPLICANT: TUBE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,847B
; FILING DATE: 26-NOV. 6582703-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BERRIDGE, WILLIAM P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 39046A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 81:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 81:
US-08-979-847B-81
```


Db 578 ISTMGPLIVLLMILLFGPCILNRLVQFVKDRISVVQALVLTQ 620

RESULT 10
US-09-309-572-14
; Sequence 14, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Moloney murine leukemia virus
; FEATURE:
; OTHER INFORMATION: env protein
US-09-309-572-14

Query Match 15.5%; Score 407.5; DB 4; Length 665;
Best Local Similarity 25.0%; Pred. No. 9.2e-33;
Matches 146; Conservative 77; Mismatches 227; Indels 133; Gaps 21;

QY 16 PFALTAPPCCCTTSS-----SPYQEFLEXRTRLPGN---IDAPSYRLSKGNSTF-- 62
Db 96 PFSSPGPPCCSGSGSPGSRDCEELTSLTPRCNTAMNRLKLDQTHKS---NEGFYV 151
QY 63 TATHMTPR---NC-----YNSATLCMHANTHYWTGK-----MINPSCPGGLGATVC-- 105
Db 152 CPQPHRPRESKSCGPDSPFYCAWGCETTCRAYWKPSSSWDFITVANNLTSDQAVQCKD 211
QY 106 -----WTFYTHTSMDSGG---GI----- 120
Db 212 NKWCNPLVIRFTDAGRRVTSWTTGHWGLRLYVSGQDGLTFGIRLRYQNLGPRVPIGPN 271
QY 121 -----QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLH-ETLRTHRLVSLFNTLT 173
Db 272 PVLADQOPLSKPKPKVKS-----SVTKPPSGTPLSPQLPPAGTENLLNLVDGAYQ 323
QY 174 RLHEVSAQNPTNCWMLPLHFRPYISIPVPEQWNNFSTEINTTSLVGLVSNLEIT--- 230
Db 324 ALNLTSPDKTQECWLCVAGPPYEGVAVLGTYSNHTSAPANCVASOHLKTLSEVTGQG 383
QY 231 -----HTSNLTCVKFSNTIDTSSOCIRWVTPPTRIIVCLPSGIPFVCCTSAYHCLNGS 283
Db 384 LCIGAVPKTHQALC---NTQTSSRGSYYLVA-----PTGTWACSTGLTPCISTT 431
QY 284 -----SESIMCFLSFLVPPMTIYTDQDLYNHVVPKHNKRVPI---LPFVIRAGVLRGLGTG 336
Db 432 ILNLTDDYCVLVELPRVTYVHSPSYVGLPERSNRHKREPVSUTLALLLGLTGGIAAG 491
QY 337 IGSITT---STQFYKLSQINGMEQVTDLSVTLQDQLNSLAAVLQNRALDILLTAKR 393
Db 492 IGTGTTALMATQFOQLQAAVQDRLREVEKSI SNLEKSLTSLSEVVLQNRGLDILLFKE 551
QY 394 GGTCLFLGERCYVYVQSRIVTEKVKIEIRDRIOCRABELQNTERM---GLLSQWMPW--- 447
Db 552 GGLCAALKERCCFYADHTGLVRDSMAKRLRNQOKLFESTQGWFEGLFNR-SPWFTTL 610
QY 448 VLPFLGLPAAIILLLFGPCIFNLVVKFVSSRIEAVKLQWVLQ 490
Db 611 ISTMGPLIVLLMILLFGPCILNRLVQFVKDRISVVQALVLTQ 653

RESULT 11
US-09-718-096-14
; Sequence 14, Application US/09718096

; Patent No. 6589763
; GENERAL INFORMATION:
; APPLICANT: Von Laer, Meike-Dorothee
; TITLE OF INVENTION: RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV
; FILE REFERENCE: 35-195
; CURRENT APPLICATION NUMBER: US/09/718,096
; CURRENT FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: DE 19856463.5
; PRIOR FILING DATE: 1998-11-26
; PRIOR APPLICATION NUMBER: EP 99250415.9
; PRIOR FILING DATE: 1999-11-25
; PRIOR APPLICATION NUMBER: US 09/309,572
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Moloney murine leukemia virus
; FEATURE:
; OTHER INFORMATION: env protein
US-09-718-096-14

Query Match 15.5%; Score 407.5; DB 4; Length 665;
Best Local Similarity 25.0%; Pred. No. 9.2e-33;
Matches 146; Conservative 77; Mismatches 227; Indels 133; Gaps 21;

QY 16 PFALTAPPCCCTTSS-----SPYQEFLEXRTRLPGN---IDAPSYRLSKGNSTF-- 62
Db 96 PFSSPGPPCCSGSGSPGSRDCEELTSLTPRCNTAMNRLKLDQTHKS---NEGFYV 151
QY 63 TATHMTPR---NC-----YNSATLCMHANTHYWTGK-----MINPSCPGGLGATVC-- 105
Db 152 CPQPHRPRESKSCGPDSPFYCAWGCETTCRAYWKPSSSWDFITVANNLTSDQAVQCKD 211
QY 106 -----WTFYTHTSMDSGG---GI----- 120
Db 212 NKWCNPLVIRFTDAGRRVTSWTTGHWGLRLYVSGQDGLTFGIRLRYQNLGPRVPIGPN 271
QY 121 -----QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLH-ETLRTHRLVSLFNTLT 173
Db 272 PVLADQOPLSKPKPKVKS-----SVTKPPSGTPLSPQLPPAGTENLLNLVDGAYQ 323
QY 174 RLHEVSAQNPTNCWMLPLHFRPYISIPVPEQWNNFSTEINTTSLVGLVSNLEIT--- 230
Db 324 ALNLTSPDKTQECWLCVAGPPYEGVAVLGTYSNHTSAPANCVASOHLKTLSEVTGQG 383
QY 231 -----HTSNLTCVKFSNTIDTSSOCIRWVTPPTRIIVCLPSGIPFVCCTSAYHCLNGS 283
Db 384 LCIGAVPKTHQALC---NTQTSSRGSYYLVA-----PTGTWACSTGLTPCISTT 431
QY 284 -----SESIMCFLSFLVPPMTIYTDQDLYNHVVPKHNKRVPI---LPFVIRAGVLRGLGTG 336
Db 432 ILNLTDDYCVLVELPRVTYVHSPSYVGLPERSNRHKREPVSUTLALLLGLTGGIAAG 491
QY 337 IGSITT---STQFYKLSQINGMEQVTDLSVTLQDQLNSLAAVLQNRALDILLTAKR 393
Db 492 IGTGTTALMATQFOQLQAAVQDRLREVEKSI SNLEKSLTSLSEVVLQNRGLDILLFKE 551
QY 394 GGTCLFLGERCYVYVQSRIVTEKVKIEIRDRIOCRABELQNTERM---GLLSQWMPW--- 447
Db 552 GGLCAALKERCCFYADHTGLVRDSMAKRLRNQOKLFESTQGWFEGLFNR-SPWFTTL 610
QY 448 VLPFLGLPAAIILLLFGPCIFNLVVKFVSSRIEAVKLQWVLQ 490
Db 611 ISTMGPLIVLLMILLFGPCILNRLVQFVKDRISVVQALVLTQ 653

RESULT 12
US-09-554-572-26
; Sequence 26, Application US/09554572
; Patent No. 6573091
; GENERAL INFORMATION:

APPLICANT: NATURE TECHNOLOGY, INC.
TITLE OF INVENTION: CHIMERIC VIRAL PACKAGING SIGNAL WITHOUT CAG GENE
FILE REFERENCE: 228.00030201
CURRENT APPLICATION NUMBER: US/09/554.572
CURRENT FILING DATE: 2000-09-18
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26
LENGTH: 1312
TYPE: PRT
ORGANISM: Murine leukemia virus
US-09-554-572-26

Query Match 15.5%; Score 407.5; DB 4; Length 1312;
Best Local Similarity 25.0%; Pred. No. 2.7e-32;
Matches 146; Conservative 77; Mismatches 227; Indels 133; Gaps 21;

QY 16 PFALTAPPCCTTSS-----SPYQFLXRTLPGN---IDAPSYRSLKGNSTP-- 62
DB 743 PFSPPGPPCCSGSGSCDCEPLTSLTPRCNTAMNRLKLDQTHKS-----NEGFYV 798
QY 63 TATHMPR--NC-----YNSATLCMHANTHYWTGK-----MINPSCPGGLGATVC-- 105
DB 799 CPGHRPRESKSCGPPDFCAWGCETTGAYWKPSSWDFITVNNLTSQAVQCKD 858
QY 106 -----WTFHTMSDGG--GI----- 120
DB 859 NKWCNPLVIRFTDAGRVRVTSWTTGHYGLRYVSGDPGLTFGIRLRYQNLGPRVPIGN 918
QY 121 -----QOAREKQVKAISQLRCHSTPSPYKGLVLKSLH-ETLRTHRLVLSLFWTLT 173
DB 919 PVLADQQPLSKPKPVKSP-----SVTKPPSGTPLSPQLPPAGTENLRLNLDVGAYQ 970
QY 174 RLHEVSAQNPNTNCWMLPLHFRPPIYISIPVPEQWNNFSTEINTTSLVGLVLSNLEIT-- 230
DB 971 ALNLTSPDKTQECWLCVLVAGPPYEGVAVLGTYSNHTSAPANCVASOHLKTLTSEVTGQG 1030
QY 231 -----HTSNLTCVKFSNTIDTTSQCIRWVTPPTTRIVCLPSGIFVCGTSAYHCLNGS 283
DB 1031 LCIGAVPKTHQALC---NTTQTSRSGSYLVA-----PTGTMACSTGLTPCISTT 1078
QY 284 -----SESNCFLSLVPPMTIYTEQDLNHHVVPKHNKRVPI---LPFVIRAGVLGRGTG 336
DB 1079 ILNLTDDCYVLVELWPRVTHSPSYVYGLFERNRHKRPVSLTALLGLGLTMGGTAAG 1138
QY 337 IGSTTT---STOFYKLSQEIINGMEQVDSLVTLQDQNSLAADVLRALDLTAKR 393
DB 1139 IGTGTALMATQQQQQAQVDDLEVEKSIKSLTSLSEVVLQNRRLGDLLEPLKE 1198
QY 394 GGTCLFLGEBRCYVYVNSQRIIVTEKVEIRDRIOCRABEELQNTWR---GLLSQWMPW--- 447
DB 1199 GGLCAALKERCCFYADHTGLVRDSMAKRLRNLQOKLPSTQCFEGLFNR-SPWFTTL 1257
QY 448 VLPPLGALAILLLLLFGPCIFNLLVKFVSSRIEAVKQWVLO 490
DB 1258 ISTIMGLPILVLLMILLFGPCILNLVGVFKDRISVQALVLTQ 1300

RESULT 13
US-09-075-272-5
; Sequence 5, Application US/09075272
; Patent No. 6136598
; GENERAL INFORMATION:
; APPLICANT: MILLER, A. DUSTY
; APPLICANT: WOLGAMOT, GREG
; APPLICANT: BONHAM, LYNN
; TITLE OF INVENTION: MUS DUNNI ENDOGENOUS RETROVIRAL
; TITLE OF INVENTION: PACKAGING CELL LINES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco
STATE: California
COUNTRY: U.S.A.
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/075,272
FILING DATE: 08-MAY-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/046,140
FILING DATE: 09-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Poor, Brian W.
REGISTRATION NUMBER: 32,928
REFERENCE/DOCKET NUMBER: 14538A-003710
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 467-9600
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 673 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-075-272-5

Query Match 14.8%; Score 389.5; DB 3; Length 673;
Best Local Similarity 27.7%; Pred. No. 6.7e-31;
Matches 106; Conservative 62; Mismatches 130; Indels 85; Gaps 13;

QY 158 LRTHRLVLSLFWTLRLHEVSAQNPNTNCWMLPLHFRPPIYISIPVPEQWNNFSTEINTT 217
DB 317 LDTENRLVLSVQAGFLVLRNTNMTQSCWLC-----YASNP----- 353
QY 218 VLVGPLVLSNLEITHTSNLT-----CV-----KFSNTIDTTSQCI----- 252
DB 354 -----PYEGIAQTRTYNTSDHSCQCLWGENRKLTLTAVSNGCLCGQVPODKWHLNQTQ 409
QY 253 -----RWTPPTTRIVCLPSGIFVCGTSAYHCLN-----GSSESNCFLSLVPPMTIY 300
DB 410 NIRPNKGGQVLVPPIDTV-----WACNTGLTFCISMSVFNSSKDFCILVQLIPRLLYH 462
QY 301 TEQDLNHHVVPKHNKRVPI---LPFVIRAGVLGR-LGTGIGSITTSQFYKLSQEIING 356
DB 463 DDSFLDKFHRVVRWKREPITLTLAVLLGLGVAAGVGTTAALIQTPRYFEELRTAMDT 522
QY 357 DMEQVTDLSVTLQDQNSLAADVLRALDLTAKRGGTCLFLGEBRCYVYVNSQRIIVTE 416
DB 523 DLRAIHSITKLEESITLSSEVVLQNRRLGDLFLKEGGLCAALKECCFYVDHSQVVKD 582
QY 417 KVKEIRDRIOCRABEELQNTWRGLLSQW---PWLPFL-----GPLAAILLLLLFGPCIF 469
DB 583 SMAKRLERLDIRQRESKQGM--FESWFKSPWLTLLSTIAGPLIILLLLLLTFGPCIIL 640
QY 470 NLLVKFVSSRIEAVKQWVLOME 492
DB 641 NKLVAFIRERINAVQV-MVLKQQ 662

RESULT 14
US-09-315-127-11
; Sequence 11, Application US/09315127
; Patent No. 6448390
; GENERAL INFORMATION:
; APPLICANT: The University of Tennessee, c/o Richard Cox
; TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and
; TITLE OF INVENTION: Liposome Vectors and Use in Gene and Drug Therapy
; FILE REFERENCE: 44137-5023, U. of Tennessee


```
Qy 488 VLQ 490
Db 631 LTQ 633

RESULT 17
US-09-315-127-9
; Sequence 9, Application US/09315127
; Patent No. 648390
; GENERAL INFORMATION:
; APPLICANT: The University of Tennessee, c/o Richard Cox
; TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and
; FILE REFERENCE: 44137-5023, U. of Tennessee
; CURRENT APPLICATION NUMBER: US/09/315.127
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQ. ID NO.
; OTHER INFORMATION: 6, envelope protein produced by retroviral vector
; OTHER INFORMATION: of seq. id no. 5
US-09-315-127-9

Query Match 14.3%; Score 376.5; DB 4; Length 645;
Best Local Similarity 28.4%; Pred. No. 1.4e-29;
Matches 103; Conservative 63; Mismatches 146; Indels 51; Gaps 10;

Qy 160 THRLVSLFNTTLRLHEVSAQNPTNCWMLPLHFRPYISIPVEQNNFST----- 211
Db 290 TGDRLNLVEGAYQALNLTNPDKTQECWCLVSGPPYVGVVGVTVTNHSTAPASCTAT 349
Qy 212 -----EINTSVLVGPLVSNLEITHTSNLTVCVKSNIIDTSSQICRWVPTTRIVC 263
Db 350 SQHKLTLSVETGQGLNGALPK-----THQA-----LCNTTQSAGS-----GSYLA 391
Qy 264 LPSGIFVCGTSAYHCLNGS-----SESMCFSLFVLPPMTIYTEQDLNVHVVPKHNKRV 319
Db 392 APAGTMWACSTGLTPCLSTTNLNTTDCYLVLEWPRIIYHSPDMYVGOLEQRTKYREP 451
Qy 320 I---LPFVIRAGVLRLGTIGSGITTS---TQFYKLSQINGMEQVTDLSVLTLODQLN 373
Db 452 VSLTALLLGLTWGTAAGIGTGTALIKTQOPEQLHAAIQTDLNEVEKSIITNLEKSLT 511
Qy 374 SLAAVVLQNRALDLLTAKRGCTCLFLGEBRCYVNSRIVTEKVKIRDIQCRABEELQ 433
Db 512 SLSEVVLQNRRLDGLFLKEGGLCAALKECCFYADHTGLVRDSMAKLRERLNQRKLFE 571
Qy 434 NTERW---GLLSQMPW---VLPFLGPLAAILLLFGPCIFNLLVKFVSSRIEAVKLQ 487
Db 572 SGQWFGQGNR-SPWFTTLSTINGPLVILLVLLLLFGPCILNRLVQFVKDRISVQVALV 630
Qy 488 VLQ 490
Db 631 LTQ 633

RESULT 18
US-09-111-085-2
; Sequence 2, Application US/09111085
; Patent No. 610034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111.085
; CURRENT FILING DATE: 1998-07-07

; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Porcine retrovirus
US-09-111-085-2

Query Match 14.3%; Score 376.5; DB 3; Length 660;
Best Local Similarity 27.5%; Pred. No. 1.4e-29;
Matches 111; Conservative 68; Mismatches 153; Indels 71; Gaps 13;

Qy 135 QLTRGHSTPSPY---KGLVLSKLHETLRLVSLNLENTLRLHEVSAQNPTNCWMLCP 191
Db 272 QEORPSPSDYNTTSGVPTPEPNTIKTKGAKFLSLIQGAFOALNSTTPEATSSCWLC 331
Qy 192 LHERPYISIPVPEQNNFSTEINTSVLVGPLVSNLEITHTSNLTVCV-----KFSNTIDT 246
Db 332 SGPPYEGMARGGKFN-----VTKEHRDQCTWGSQNKULTLTVS 370
Qy 247 TSSQCIRWVTPTRIVCLPISGIF-----FVCGTSAYHCLN-----GSSES 286
Db 371 GKGTGCMVPPSHQHLCHNTEAFNRTSQYLVPGYDRWACNTGLTPCVSTLVFNQTKD 430
Qy 287 MCFSLFVLPPMTIYTEQ---DLNVHVVPKHNKRVPI---LPFVIRAGVLRLGTIGSI 340
Db 431 FCVMVQIVPRVYVPEKAVLDEYDYRNRP--KREPISLTLAVMLGLGVAAGVGTGTAAL 488
Qy 341 TTSTQFYK---LSQINGMEQVTDLSVLTLODQLNSLAAVLQNRALDLLTAKRGCT 396
Db 489 ITGPQOLEKGLSLNLRIVTEDLQALEKSVSNLEESLTSLSSEVVLQNRRLDGLFLKEGGL 548
Qy 397 CLFLGEBRCYVNSRIVTEKVKIRDIQCRABEELQNRERERRERADQGW---PWVLPEL- 452
Db 549 CVALKKECCFYVDSHSGNAIRDSMKLRERLERERRERADQGW---PWVLPEL- 452
Qy 453 ---GPLAAILLLFGPCIFNLLVKFVSSRIEAVKLQVLOME 492
Db 607 ALTGPLVLLLLTVGPCILNRFVAFVRERVSQVI-MVLRQQ 648

RESULT 19
US-09-376-781-5
; Sequence 5, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papiya T.
; APPLICANT: Patience, Clive
; APPLICANT: Anderson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PERVA
; OTHER INFORMATION: polypeptide sequence taken from GenBank Accession
; OTHER INFORMATION: No. 6261806 Y12238 for comparison.
US-09-376-781-5

Query Match 14.3%; Score 376.5; DB 3; Length 660;
Best Local Similarity 27.5%; Pred. No. 1.4e-29;
```



```
; MOLECULE TYPE: protein
US-08-486-099-107

Query Match
Best Local Similarity 39.8%; Score 364; DB 3; Length 192;
Matches 68; Conservative 44; Mismatches 59; Indels 0; Gaps 0;

QY 318 VPILPFVIRAGVLGRGTGSGITTTQFYKLSQEINGMEQVTSVLTLQDLNSLAA 377
Db 2 IQILPLFVGLGITTAVSTGAAGLVSITQYTKLSHQLISDVQAISSTIQDLQDVSLAE 61

QY 378 VVLQNRRLDLLTAKRGCTCLFLGEERCYVYNOSRIVTEKVKIRDRIOCRAEELQNTER 437
Db 62 VVLQNRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKNLQDLERRRLQIDNPF 121

QY 438 WGLLSQMPVLPFLGPLAALILLFLGPCIFNLLVKFVSSRIEAVKLQMV 488
Db 122 WTSFHGFLPYVMPPLGLLCLLVLSFGPIIFNKLMTFIKHQIESIQAKPI 172

RESULT 22
US-08-360-107A-117
; Sequence 117, Application US/08360107A
; Patent No. 6017536
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 149
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,107A
; FILING DATE: 20-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 117:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-360-107A-117

Query Match
Best Local Similarity 39.8%; Score 364; DB 3; Length 192;
Matches 68; Conservative 44; Mismatches 59; Indels 0; Gaps 0;

QY 318 VPILPFVIRAGVLGRGTGSGITTTQFYKLSQEINGMEQVTSVLTLQDLNSLAA 377
Db 2 IQILPLFVGLGITTAVSTGAAGLVSITQYTKLSHQLISDVQAISSTIQDLQDVSLAE 61

QY 378 VVLQNRRLDLLTAKRGCTCLFLGEERCYVYNOSRIVTEKVKIRDRIOCRAEELQNTER 437
Db 62 VVLQNRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKNLQDLERRRLQIDNPF 121

QY 438 WGLLSQMPVLPFLGPLAALILLFLGPCIFNLLVKFVSSRIEAVKLQMV 488
Db 122 WTSFHGFLPYVMPPLGLLCLLVLSFGPIIFNKLMTFIKHQIESIQAKPI 172

RESULT 23
US-08-484-223B-107
; Sequence 107, Application US/08484223B
; Patent No. 6020459
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 245
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,223B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-484-223B-107

Query Match
Best Local Similarity 39.8%; Score 364; DB 3; Length 192;
Matches 68; Conservative 44; Mismatches 59; Indels 0; Gaps 0;

QY 318 VPILPFVIRAGVLGRGTGSGITTTQFYKLSQEINGMEQVTSVLTLQDLNSLAA 377
Db 2 IQILPLFVGLGITTAVSTGAAGLVSITQYTKLSHQLISDVQAISSTIQDLQDVSLAE 61

QY 378 VVLQNRRLDLLTAKRGCTCLFLGEERCYVYNOSRIVTEKVKIRDRIOCRAEELQNTER 437
Db 62 VVLQNRGLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKNLQDLERRRLQIDNPF 121
```

QY 438 WGLLSQWMPWVLPFLGPAALILLLLFGPCIFNLLVKFVSSRIEAVKLOWV 488
Db 122 WTSFHGFLPYVMPFLGPLLCLLVLSFGPIIFNKLMTFIKHQIESIQAKPI 172

RESULT 24
US-08-919-597-107
; Sequence 107, Application US/08919597
; Patent No. 6054265
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/470,896
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-919-597-107

Query Match 13.8%; Score 364; DB 3; Length 192;
Best Local Similarity 39.8%; Pred. No. 3.9e-29;
Matches 68; Conservative 44; Mismatches 59; Indels 0; Gaps 0;
QY 318 VPILPFVIRAGVLGRGTGIGSITTSQFYVKLSQEIINGMEQVTSVLTLQDQLNSLAA 377
Db 2 IQILPLFVLGITTAVSTGAAGLVSTQYTKLSHQLISDVQAISSTIQDLQDQVDSLAE 61
QY 378 VVLQNRALDILLAKRGCTCLFLGEECCYVNSRIVTEKVKIIRDIQCRABELQNTER 437
Db 62 VVLQNRGLDILLTAEOGGICLQKCCFYANKSGIVRDKIKNLQDQLERRRRQLIDNPF 121
QY 438 WGLLSQWMPWVLPFLGPAALILLLLFGPCIFNLLVKFVSSRIEAVKLOWV 488
Db 122 WTSFHGFLPYVMPFLGPLLCLLVLSFGPIIFNKLMTFIKHQIESIQAKPI 172

RESULT 25
US-08-475-668A-107
; Sequence 107, Application US/08475668A
; Patent No. 6060065
; GENERAL INFORMATION:
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 211
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,668A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-475-668A-107

Query Match 13.8%; Score 364; DB 3; Length 192;
Best Local Similarity 39.8%; Pred. No. 3.9e-29;
Matches 68; Conservative 44; Mismatches 59; Indels 0; Gaps 0;
QY 318 VPILPFVIRAGVLGRGTGIGSITTSQFYVKLSQEIINGMEQVTSVLTLQDQLNSLAA 377
Db 2 IQILPLFVLGITTAVSTGAAGLVSTQYTKLSHQLISDVQAISSTIQDLQDQVDSLAE 61
QY 378 VVLQNRALDILLAKRGCTCLFLGEECCYVNSRIVTEKVKIIRDIQCRABELQNTER 437
Db 62 VVLQNRGLDILLTAEOGGICLQKCCFYANKSGIVRDKIKNLQDQLERRRRQLIDNPF 121
QY 438 WGLLSQWMPWVLPFLGPAALILLLLFGPCIFNLLVKFVSSRIEAVKLOWV 488
Db 122 WTSFHGFLPYVMPFLGPLLCLLVLSFGPIIFNKLMTFIKHQIESIQAKPI 172

RESULT 26
US-08-485-551A-107
; Sequence 107, Application US/08485551A
; Patent No. 6068973
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.

```
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 211
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,551A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-485-551A-107

Query Match 13.8%; Score 364; DB 3; Length 192;
Best Local Similarity 39.8%; Pred. No. 3.9e-29;
Matches 68; Conservative 44; Mismatches 59; Indels 0; Gaps 0;

QY 318 VPILPFVIRAGVLGRGTGIGSTTTQFYKLSQEIINGMEQVTSVLTLQDQLNSLAA 377
Db 2 IQLPLFVGLGITAVSTGAAGLGSITQYTKLSHQLISDVQAISSTIQDLQDQVDSLAE 61
QY 378 VVLQNRALDLLAKRGTCFLGEECCYVYVNSRIVTEKVKIRDRIOCRABEQLQNTER 437
Db 62 VVLQNRRLGLDLTAEOGGICLALQEKCCFYANKSGIVRDKIKNLQDQLERRRRQLIDNPF 121
QY 438 WGLLSQMPVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLMV 488
Db 122 WTSFHGFLPYVMPLLGPLLCLLLVLSFGPIIFNKLMTFIKHQIESIQAKPI 172

RESULT 27
US-08-471-913A-107
; Sequence 107, Application US/08471913A
; Patent No. 6093794
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 214
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,913A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-471-913A-107

Query Match 13.8%; Score 364; DB 3; Length 192;
Best Local Similarity 39.8%; Pred. No. 3.9e-29;
Matches 68; Conservative 44; Mismatches 59; Indels 0; Gaps 0;

QY 318 VPILPFVIRAGVLGRGTGIGSTTTQFYKLSQEIINGMEQVTSVLTLQDQLNSLAA 377
Db 2 IQLPLFVGLGITAVSTGAAGLGSITQYTKLSHQLISDVQAISSTIQDLQDQVDSLAE 61
QY 378 VVLQNRALDLLAKRGTCFLGEECCYVYVNSRIVTEKVKIRDRIOCRABEQLQNTER 437
Db 62 VVLQNRRLGLDLTAEOGGICLALQEKCCFYANKSGIVRDKIKNLQDQLERRRRQLIDNPF 121
QY 438 WGLLSQMPVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLMV 488
Db 122 WTSFHGFLPYVMPLLGPLLCLLLVLSFGPIIFNKLMTFIKHQIESIQAKPI 172

RESULT 28
US-08-485-264A-107
; Sequence 107, Application US/08485264A
; Patent No. 6228983
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING
; TITLE OF INVENTION: RESPIRATORY SYNCYTIAL VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
```

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,264A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-021
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-485-264A-107

Query Match      13.8%; Score 364; DB 3; Length 192;
Best Local Similarity 39.8%; Pred. No. 3.9e-29;
Matches 68; Conservative 44; Mismatches 59; Indels 0; Gaps 0;

QY 318 VPILPFVIRAGVLGRGTGTTTQFYKLSQEIINGMEQVTSVLTQDQNSLAA 377
Db 2 IQILPLFVGLGITTAVSTGAAGLGVSTQYTKLSHQLISDVQAISSTIQDQVDSLAE 61

QY 378 VVLQNRRLDLTAKGGTCLFLGEBRCYVNCOSRIVTEKVKIRDIQCRABELQNTER 437
Db 62 VVLQNRRLDLTAKGGTCLFLGEBRCYVNCOSRIVTEKVKIRDIQCRABELQNTER 437

QY 438 WGLLSQMPWLPFLGPLAALILLFPGCIFNLLVKFVSSRIEAVKLQMV 488
Db 122 WTSFHGFLPYVWPLGFLCLLVLSFGPIFNKLMFTFKHQIESIQAKPI 172

RESULT 29
US-08-474-349A-107
; Sequence 107, Application US/08474349A
; Patent No. 633395
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING HUMAN PARAINFLUENZA
; TITLE OF INVENTION: VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 517
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,349A
```

```
;
; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-024
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-474-349A-107

Query Match      13.8%; Score 364; DB 3; Length 192;
Best Local Similarity 39.8%; Pred. No. 3.9e-29;
Matches 68; Conservative 44; Mismatches 59; Indels 0; Gaps 0;

QY 318 VPILPFVIRAGVLGRGTGTTTQFYKLSQEIINGMEQVTSVLTQDQNSLAA 377
Db 2 IQILPLFVGLGITTAVSTGAAGLGVSTQYTKLSHQLISDVQAISSTIQDQVDSLAE 61

QY 378 VVLQNRRLDLTAKGGTCLFLGEBRCYVNCOSRIVTEKVKIRDIQCRABELQNTER 437
Db 62 VVLQNRRLDLTAKGGTCLFLGEBRCYVNCOSRIVTEKVKIRDIQCRABELQNTER 437

QY 438 WGLLSQMPWLPFLGPLAALILLFPGCIFNLLVKFVSSRIEAVKLQMV 488
Db 122 WTSFHGFLPYVWPLGFLCLLVLSFGPIFNKLMFTFKHQIESIQAKPI 172

RESULT 30
US-08-470-896-107
; Sequence 107, Application US/08470896
; Patent No. 6479055
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,896
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-020
; TELECOMMUNICATION INFORMATION:
```

```
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-470-896-107

Query Match      13.8%; Score 364; DB 4; Length 192;
Best Local Similarity 39.8%; Pred. No. 3.9e-29;
Matches 68; Conservative 44; Mismatches 59; Indels 0; Gaps 0;

QY 318 VPILPFVIRAGVLGRIGTIGSITTSQFYFKLSQEIINGMEQVTDLSLVTLDQLNSLAA 377
Db 2 IQLPLFVGLGITTAVSTGAAGLGVISITQYTKLSHQLISDVQAISSTIQDLQDQVDSLAE 61

QY 378 VVLQNRRLDLLTAKRGGTCLFLGHERCYVYNOSRIVTEKVKIQRDIQCRABELQNTER 437
Db 62 VVLQNRRLDLLTAEQGGICLALQEKCCFYANKSGIVRDKIKNLQDDLERRRRLIDNPF 121

QY 438 WGLLSQMPWVLPFLGFLAALILLLLFGPCIFENLLVKFVSSRIEAVKIQMV 488
Db 122 WTSFHGFLPYVMPFLGFLCLLLVLSFGPILFNKLMTFIKHQIESIQAKPI 172
```

Search completed: January 22, 2005, 22:12:41
Job time : 48 secs

This page blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2005, 22:07:21 ; Search time 106 Seconds

(without alignments)
1680.336 Million cell updates/sec

Title: US-09-319-156B-10

Perfect score: 2634

Sequence: 1 MALPHTFTLVLLPPFALT.....KFVSSRIEAVKLMQLQMP 493

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters: 1608061

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2632	99.9	493	8 US-08-979-847-106	Sequence 106, App
2	2632	99.9	493	14 US-10-114-104-106	Sequence 106, App
3	2447	92.9	542	15 US-10-416-642-2	Sequence 2, Appli
4	2419	91.8	503	15 US-10-637-565-20	Sequence 20, Appl
5	2319	88.0	538	9 US-09-902-535-2	Sequence 2, Appli
6	2319	88.0	538	13 US-10-114-893-135	Sequence 135, App
7	2319	88.0	538	14 US-10-016-249-4	Sequence 4, Appli
8	2314	87.9	538	15 US-10-133-036-2	Sequence 2, Appli
9	2285	86.8	538	15 US-10-133-036-21	Sequence 21, Appl
10	2237	84.9	538	15 US-10-133-036-20	Sequence 20, Appl
11	1235.5	46.9	263	9 US-09-864-761-37656	Sequence 37656, A
12	1122	42.6	531	15 US-10-363-616-473	Sequence 473, App
13	591	22.4	120	15 US-10-133-036-4	Sequence 4, Appli

14	516	19.6	162	8 US-08-979-847-109	Sequence 109, App
15	516	19.6	162	14 US-10-114-104-109	Sequence 109, App
16	514.5	19.5	576	10 US-09-793-451-4	Sequence 4, Appli
17	514.5	19.5	576	14 US-10-283-722-4	Sequence 4, Appli
18	514.5	19.5	576	14 US-10-283-903-4	Sequence 4, Appli
19	480.5	18.2	469	13 US-10-001-835-168	Sequence 168, App
20	452	17.2	433	8 US-08-979-847-81	Sequence 81, Appl
21	452	17.2	433	14 US-10-430-442-87	Sequence 87, Appl
22	452	17.2	433	14 US-10-114-104-81	Sequence 81, Appl
23	423	16.1	253	10 US-09-764-891-2970	Sequence 2970, Ap
24	421	16.0	618	15 US-10-276-774-1749	Sequence 1749, Ap
25	393	14.9	676	9 US-09-954-983-11	Sequence 11, Appl
26	392	14.9	675	9 US-09-954-983-8	Sequence 8, Appli
27	391	14.8	672	9 US-09-954-983-4	Sequence 4, Appli
28	389	14.8	675	9 US-09-954-983-6	Sequence 6, Appli
29	386	14.7	672	9 US-09-954-983-5	Sequence 5, Appli
30	386	14.7	675	9 US-09-954-983-7	Sequence 7, Appli
31	378	14.4	638	16 US-10-441-949-32	Sequence 32, Appl
32	378	14.4	638	16 US-10-441-949-34	Sequence 34, Appl
33	378	14.4	638	16 US-10-441-949-40	Sequence 40, Appl
34	378	14.4	638	16 US-10-441-949-42	Sequence 42, Appl
35	378	14.4	638	16 US-10-441-949-45	Sequence 45, Appl
36	378	14.4	678	16 US-10-441-949-36	Sequence 36, Appl
37	378	14.4	678	16 US-10-441-949-38	Sequence 38, Appl
38	376.5	14.3	660	9 US-09-851-859A-5	Sequence 5, Appli
39	376.5	14.3	660	16 US-10-441-949-8	Sequence 8, Appli
40	376.5	14.3	660	16 US-10-441-949-10	Sequence 10, Appl
41	376.5	14.3	660	16 US-10-441-949-16	Sequence 16, Appl
42	376.5	14.3	660	16 US-10-441-949-18	Sequence 18, Appl
43	376.5	14.3	661	16 US-10-441-949-43	Sequence 43, Appl
44	374.5	14.2	653	14 US-10-029-656-4	Sequence 4, Appli
45	372.5	14.1	653	14 US-10-029-656-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-08-979-847-106

; Sequence 106, Application US/08979847

; Publication No. US20030039664A1

; GENERAL INFORMATION:

; APPLICANT: PERRON, HERVE

; APPLICANT: BESEME, FREDERIC

; APPLICANT: BEDIN, FREDERIC

; APPLICANT: PARANHOS-BACCALA, GLAUCIA

; APPLICANT: KOMURIAN-PRADEL, FLORENCE

; APPLICANT: JOLIVET-REYNAUD, COLETTE

; APPLICANT: MANDRAND, BERNARD

; APPLICANT: GARSON, JEREMY

; APPLICANT: TUKE, PHILIP

; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT

; TITLE OF INVENTION: THERAPEUTIC PURPOSES

; NUMBER OF SEQUENCES: 210

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OLIFF & BERRIDGE, PLC

; STREET: P.O. BOX 19928

; CITY: ALEXANDRIA

; STATE: VA

; COUNTRY: USA

; ZIP: 22320

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/979,847

; FILING DATE: 26-NOV-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: BERRIDGE, WILLIAM P.

REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 106:

SEQUENCE CHARACTERISTICS:

LENGTH: 493 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-979-847-106

Query Match 99.9%; Score 2632; DB 8; Length 493;
Best Local Similarity 100.0%; Pred. No. 5.4e-240; Indels 0; Gaps 0;
Matches 493; Conservative 0; Mismatches 0;

QY	1	MALPYHTFTLVLLPPFALTAPPCCCTTSSSPYQBFLEXRTRLPGNIDAPSYRSLSKGNS	60
DB	1	MALPYHTFTLVLLPPFALTAPPCCCTTSSSPYQBFLEXRTRLPGNIDAPSYRSLSKGNS	60
QY	61	TFTAHTMPRNCVNSATLCMHANTHYWTGKMINPSCPGGLGATVCWYTFHTSMDSGGI	120
DB	61	TFTAHTMPRNCVNSATLCMHANTHYWTGKMINPSCPGGLGATVCWYTFHTSMDSGGI	120
QY	121	QGOAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRTHTRLVSLFNTTLRLHEVSA	180
DB	121	QGOAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRTHTRLVSLFNTTLRLHEVSA	180
QY	181	QNPNCNMCPLPHFRPYISIPVPEQWNNFSTEINTTSLVGLPLVSLNLEITHTSNLTCVKF	240
DB	181	QNPNCNMCPLPHFRPYISIPVPEQWNNFSTEINTTSLVGLPLVSLNLEITHTSNLTCVKF	240
QY	241	SNITDITSSQICRWVTPTRIVCLPSGIFVCGTSAYHCLNGSSEMCFLSFLVPPMTIY	300
DB	241	SNITDITSSQICRWVTPTRIVCLPSGIFVCGTSAYHCLNGSSEMCFLSFLVPPMTIY	300
QY	301	TEODLYNHVVPKPHNKRVPILPFVIRAGVLGRGTGIGSTTTQFYKLSQINGDMEQ	360
DB	301	TEODLYNHVVPKPHNKRVPILPFVIRAGVLGRGTGIGSTTTQFYKLSQINGDMEQ	360
QY	361	VTDSLVTLODQLNSLAQVQLNRRALDLTAKGGTCLFLGECRCYVQNSRIVTEKVE	420
DB	361	VTDSLVTLODQLNSLAQVQLNRRALDLTAKGGTCLFLGECRCYVQNSRIVTEKVE	420
QY	421	IRDRIQCRABELQNTERWGLLSQMPWPVLPFLGFLAALILLLLFGPCIFNLLKVFSSRI	480
DB	421	IRDRIQCRABELQNTERWGLLSQMPWPVLPFLGFLAALILLLLFGPCIFNLLKVFSSRI	480
QY	481	EAVKLQWLQMEP 493	
DB	481	EAVKLQWLQMEP 493	

RESULT 2

US-10-114-104-106

Sequence 106, Application US/10114104

Publication No. US20030198647A1

GENERAL INFORMATION:

APPLICANT: PERRON, HERVE

BESIME, FREDERIC

BEDIN, FREDERIC

PARANHOS-BACCALA, GLAUCIA

KOMURIAN-PRADEL, FLORENCE

JOLIVET-REYNAUD, COLETTE

MANDRAND, BERNARD

GARSON, JEREMY

TUKE, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL

THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210

CORRESPONDENCE ADDRESS:

ADDRESSEE: OLIFF & BERRIDGE, PLC

STREET: P.O. BOX 19928

CITY: ALEXANDRIA

STATE: VA

COUNTRY: USA

ZIP: 22320

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patencin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/114,104

FILING DATE: 03-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/979,847

FILING DATE: 26-NOV-1997

ATTORNEY/AGENT INFORMATION:

NAME: BERRIDGE, WILLIAM P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: WPB 39046A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 106:

SEQUENCE CHARACTERISTICS:

LENGTH: 493 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 106:

US-10-114-104-106

Query Match 99.9%; Score 2632; DB 14; Length 493;
Best Local Similarity 100.0%; Pred. No. 5.4e-240;
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MALPYHTFTLVLLPPFALTAPPCCCTTSSSPYQBFLEXRTRLPGNIDAPSYRSLSKGNS	60
DB	1	MALPYHTFTLVLLPPFALTAPPCCCTTSSSPYQBFLEXRTRLPGNIDAPSYRSLSKGNS	60
QY	61	TFTAHTMPRNCVNSATLCMHANTHYWTGKMINPSCPGGLGATVCWYTFHTSMDSGGI	120
DB	61	TFTAHTMPRNCVNSATLCMHANTHYWTGKMINPSCPGGLGATVCWYTFHTSMDSGGI	120
QY	121	QGOAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRTHTRLVSLFNTTLRLHEVSA	180
DB	121	QGOAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRTHTRLVSLFNTTLRLHEVSA	180
QY	181	QNPNCNMCPLPHFRPYISIPVPEQWNNFSTEINTTSLVGLPLVSLNLEITHTSNLTCVKF	240
DB	181	QNPNCNMCPLPHFRPYISIPVPEQWNNFSTEINTTSLVGLPLVSLNLEITHTSNLTCVKF	240
QY	241	SNITDITSSQICRWVTPTRIVCLPSGIFVCGTSAYHCLNGSSEMCFLSFLVPPMTIY	300
DB	241	SNITDITSSQICRWVTPTRIVCLPSGIFVCGTSAYHCLNGSSEMCFLSFLVPPMTIY	300
QY	301	TEODLYNHVVPKPHNKRVPILPFVIRAGVLGRGTGIGSTTTQFYKLSQINGDMEQ	360
DB	301	TEODLYNHVVPKPHNKRVPILPFVIRAGVLGRGTGIGSTTTQFYKLSQINGDMEQ	360
QY	361	VTDSLVTLODQLNSLAQVQLNRRALDLTAKGGTCLFLGECRCYVQNSRIVTEKVE	420
DB	361	VTDSLVTLODQLNSLAQVQLNRRALDLTAKGGTCLFLGECRCYVQNSRIVTEKVE	420
QY	421	IRDRIQCRABELQNTERWGLLSQMPWPVLPFLGFLAALILLLLFGPCIFNLLKVFSSRI	480
DB	421	IRDRIQCRABELQNTERWGLLSQMPWPVLPFLGFLAALILLLLFGPCIFNLLKVFSSRI	480
QY	481	EAVKLQWLQMEP 493	

Db 481 EAVKQWVLMQEP 493
|||||
RESULT 3
US-10-416-642-2
; Sequence 2, Application US/10416642
; Publication No. US20040043452A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: ARVIZU, Chandra
; TITLE OF INVENTION: EMBRYOGENESIS ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0842 PCT
; CURRENT APPLICATION NUMBER: US/10/416,642
; CURRENT FILING DATE: 2003-05-13
; PRIOR APPLICATION NUMBER: 60/249,407
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040043452A1 7477736CD1
US-10-416-642-2

Query Match 92.9%; Score 2447; DB 15; Length 542;
Best Local Similarity 93.1%; Pred. No. 2.1e-222;
Matches 459; Conservative 8; Mismatches 26; Indels 0; Gaps 0;
Qy 1 MALPYHTFLVLPFPALTAAPPCCCTSSSPYQBFLEXRTRLPGNIDAPSYSLSGNS 60
Db 1 MALPYCIFLTVLSPFSLTAPSPCHRTSSSPYQAFLEWRMRPHLDAPSYSLSGNP 60
Qy 61 TFTAHTMPRNCYSATLCMHANTHYWTGKINPSCPGGLGATVCWTYFTHTSMSGGI 120
Db 61 AFTAHTMPHNCYSATLCMHANTHYWTGKINPSCPGGLGATVCWTYFTHTSMSGGI 120
Qy 121 QGAREKQVKEAISQLTRGHSTSPYKGLVLSKHLHETLTHLVSLFNTTLRLHEVSA 180
Db 121 QGAREKQVKEAISQLTRGHSTSPYKGLVLSKHLHETLTHLVSLFNTTLRLHEVSA 180
Qy 181 QNPTNCWMLPLHFRPYISIPVPEQWNNFSTEINTTSLVGLVSNLEITHTSNLTGVKF 240
Db 181 QNPTNCWMLPLHFRPYISIPVPEQWNNFSTEINTTSLVGLVSNLEITHTSNLTGVKF 240
Qy 241 SNTIDTSSQICIRWTPPTTRIVCLPSGIFPVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
Db 241 SNTIATNTSQCIRWTPPTQIVCLPSGIFPVCGTSAYHCLNGSSESMCFLSFLVPPMTIY 300
Qy 301 TEODLYNHVVPKPHNKRVPILPFVIRAGVLRIGTGIGSTTSTQFYKLSQBIINGMEQ 360
Db 301 TEODLYNHVVPKPHNKRVPILPFVIRAGVLRIGTGIGSTTSTQFYKLSQBIINGMEQ 360
Qy 361 VTDLSVTLQDQNSLAALVQNRRLDLLTAKRGCTCLFGEERCYYVNSQSRIVTEKVKE 420
Db 361 VTDLSVTLQDQNSLAALVQNRRLDLLTAKRGCTCLFGEERCYYVNSQSRIVTEKVKE 420
Qy 421 IRDRIQCRABEELQNTERWGLLSQWMPWLPFLGPLAALILLFPGCIFNLLVKFVSSRI 480
Db 421 IRDRIQCRABEELQNTERWGLLSQWMPWLPFLGPLAALILLFPGCIFNLLVKFVSSRI 480
Qy 481 EAVKQWVLMQEP 493
Db 481 EAVKQWVLMQEP 493

RESULT 4
US-10-637-565-20
; Sequence 20, Application US/10637565

; Publication No. US20040043381A1
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glauca
; APPLICANT: PERRON, Herve
; APPLICANT: KOMURIAN-PRADEL, Florence
; TITLE OF INVENTION: THE LTR REGION OF MSRV-1 AND THE PROTEINS IT ENCODES, AND PROBES
; TITLE OF INVENTION: METHODS FOR DETECTING MSRV-1 RETROVIRUS
; FILE REFERENCE: 110257
; CURRENT APPLICATION NUMBER: US/10/637,565
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: US/09/890,340
; PRIOR FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: PCT/IB00/00159
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: EP 99420041.8
; PRIOR FILING DATE: 1999-02-15
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 20
; LENGTH: 503
; TYPE: PRT
; ORGANISM: MSRV-1 retrovirus
US-10-637-565-20

Query Match 91.8%; Score 2419; DB 15; Length 503;
Best Local Similarity 100.0%; Pred. No. 8.4e-220;
Matches 454; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 40 RTPLPGNIDAPSYRSLSGNSTTATHMPRNCYSATLCMHANTHYWTGKINPSCPGG 99
Db 1 RTPLPGNIDAPSYRSLSGNSTTATHMPRNCYSATLCMHANTHYWTGKINPSCPGG 60
Qy 100 LGATVCWTYFTHTSMSGGIGQAREKQVKEAISQLTRGHSTSPYKGLVLSKHLHETLR 159
Db 61 LGATVCWTYFTHTSMSGGIGQAREKQVKEAISQLTRGHSTSPYKGLVLSKHLHETLR 120
Qy 160 THTRLVSLFNTTLRLHEVSAQNPNTNCWMLPLHFRPYISIPVPEQWNNFSTEINTTSLV 219
Db 121 THTRLVSLFNTTLRLHEVSAQNPNTNCWMLPLHFRPYISIPVPEQWNNFSTEINTTSLV 180
Qy 220 VGPLVSNLEITHTSNLTGVKFSTIDTSSQICIRWTPPTTRIVCLPSGIFVCGTSAYHC 279
Db 181 VGPLVSNLEITHTSNLTGVKFSTIDTSSQICIRWTPPTTRIVCLPSGIFVCGTSAYHC 240
Qy 280 LINGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIRAGVLRIGTGIGS 339
Db 241 LINGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIRAGVLRIGTGIGS 300
Qy 340 ITTSTQFYKLSQBIINGMEQVTDLSVTLQDQNSLAALVQNRRLDLLTAKRGCTCLF 399
Db 301 ITTSTQFYKLSQBIINGMEQVTDLSVTLQDQNSLAALVQNRRLDLLTAKRGCTCLF 360
Qy 400 LGEERCYYVNSQSRIVTEKVKEIRDRIQCRABEELQNTERWGLLSQWMPWLPFLGPLAALI 459
Db 361 LGEERCYYVNSQSRIVTEKVKEIRDRIQCRABEELQNTERWGLLSQWMPWLPFLGPLAALI 420
Qy 460 LLLFGPCIFNLLVKFVSSRIEAVKQWVLMQEP 493
Db 421 LLLFGPCIFNLLVKFVSSRIEAVKQWVLMQEP 454

RESULT 5
US-09-902-535-2
; Sequence 2, Application US/09902535
; Patent No. US20020102530A1
; GENERAL INFORMATION:
; APPLICANT: Keith, Jr., James C.
; APPLICANT: McCoy, John M.
; APPLICANT: Mi, Sha
; TITLE OF INVENTION: Methods and compositions for diagnosing
; TITLE OF INVENTION: and treating preeclampsia and gestational trophoblast
; TITLE OF INVENTION: disorders
; FILE REFERENCE: GIN-6006B4

; CURRENT APPLICATION NUMBER: US/09/902,535
; CURRENT FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 60/216,657
; PRIOR FILING DATE: 2000-07-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-902-535-2

Query Match 88.0%; Score 2319; DB 9; Length 538;
Best Local Similarity 88.6%; Pred. No. 2.7e-210;
Matches 436; Conservative 18; Mismatches 38; Indels 0; Gaps 0;

```
QY 1 MALPYHTFTLLPPFALTAPPCCCTTSSSPYQBFELXRTLPGNIDAPSYRSLSKGNS 60
   |||||
Db 1 MALPYHIFLFTVLLPSFTLTAPPCCMTSSSPYQBFELRMQRPNGIDAPSYRSLSKGTP 60
   |||||
QY 61 TFTAHTMPRNCYNATLCHMANTHYWTGMINPSCPGGLGATVCWTYFTHTSMDSGGGI 120
   |||||
Db 61 TFTAHTMPRNCYHSATLCHMANTHYWTGMINPSCPGGLGATVCWTYFTHTSMDSGGGV 120
   |||||
QY 121 QGOAREKQVKEAISQTRGHSTSPYKGLVLSKLHETLRTHTLRLVSLFNTTLRLHEVSA 180
   |||||
Db 121 QGOAREKHVKEVISQLTRVHGTSPPYKGLDLSKLHETLRTHTLRLVSLFNTTLRLHEVSA 180
   |||||
QY 181 QNPTNCWMLPLHFRYISIPVPEOWNFSTEINTTSLVGLPLVSNLEITHTSNLTCVKF 240
   |||||
Db 181 QNPTNCWICLPLNFRYVSIIPVPEOWNFSTEINTTSLVGLPLVSNLEITHTSNLTCVKF 240
   |||||
QY 241 SNTIDTSSQICIRWVTPTRIVCLPSGIFVCGTSAYHCLNGSSESMCFSLFLVPPMTIY 300
   |||||
Db 241 SNTYTTNSQICIRWVTPTRIVCLPSGIFVCGTSAYHCLNGSSESMCFSLFLVPPMTIY 300
   |||||
QY 301 TEODLYNHVVPKPNKRVPILPFVIRAGVLGRGTGISTTSTQFYKLSQELNGDMEQ 360
   |||||
Db 301 TEODLYNVISKPRNKRVPILPFVIRAGVLGRGTGISTTSTQFYKLSQELNGDMEQ 360
   |||||
QY 361 VTDSLVTLODQNSLAAVLQNRALDLLTAKGGTCLFLGEECCYVNSGIVTEKVK 420
   |||||
Db 361 VADSLVTLODQNSLAAVLQNRALDLLTAKGGTCLFLGEECCYVNSGIVTEKVK 420
   |||||
QY 421 IRDRIQRAEELQNTERWGLLSQWMPWLPFLGFLAAILLLFGPCIFNLLVKFVSSRI 480
   |||||
Db 421 IRDRIQRAEELQNTGPWGLLSQWMPWLPFLGFLAAILLLFGPCIFNLLVNFVSSRI 480
   |||||
QY 481 EAVKLQWLQME 492
   |||||
Db 481 EAVKLQWEPKMQ 492
   |||||
```

RESULT 6

US-10-016-893-135.
; Sequence 135, Application US/10114893
; Publication No. US20020193567A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Carlin-Duckett, McKeough
; APPLICANT: Kelleher, Kerry S.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: GI 6000-10A
; CURRENT APPLICATION NUMBER: US/10/114,893

; CURRENT FILING DATE: 2002-04-02
; EARLIER APPLICATION NUMBER: 09/413,232
; NUMBER OF SEQ ID NOS: 321
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 135
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-114-893-135

Query Match 88.0%; Score 2319; DB 13; Length 538;
Best Local Similarity 88.6%; Pred. No. 2.7e-210;
Matches 436; Conservative 18; Mismatches 38; Indels 0; Gaps 0;

```
QY 1 MALPYHTFTLLPPFALTAPPCCCTTSSSPYQBFELXRTLPGNIDAPSYRSLSKGNS 60
   |||||
Db 1 MALPYHIFLFTVLLPSFTLTAPPCCMTSSSPYQBFELRMQRPNGIDAPSYRSLSKGTP 60
   |||||
QY 61 TFTAHTMPRNCYNATLCHMANTHYWTGMINPSCPGGLGATVCWTYFTHTSMDSGGGI 120
   |||||
Db 61 TFTAHTMPRNCYHSATLCHMANTHYWTGMINPSCPGGLGATVCWTYFTHTSMDSGGGV 120
   |||||
QY 121 QGOAREKQVKEAISQTRGHSTSPYKGLVLSKLHETLRTHTLRLVSLFNTTLRLHEVSA 180
   |||||
Db 121 QGOAREKHVKEVISQLTRVHGTSPPYKGLDLSKLHETLRTHTLRLVSLFNTTLRLHEVSA 180
   |||||
QY 181 QNPTNCWMLPLHFRYISIPVPEOWNFSTEINTTSLVGLPLVSNLEITHTSNLTCVKF 240
   |||||
Db 181 QNPTNCWICLPLNFRYVSIIPVPEOWNFSTEINTTSLVGLPLVSNLEITHTSNLTCVKF 240
   |||||
QY 241 SNTIDTSSQICIRWVTPTRIVCLPSGIFVCGTSAYHCLNGSSESMCFSLFLVPPMTIY 300
   |||||
Db 241 SNTYTTNSQICIRWVTPTRIVCLPSGIFVCGTSAYHCLNGSSESMCFSLFLVPPMTIY 300
   |||||
QY 301 TEODLYNHVVPKPNKRVPILPFVIRAGVLGRGTGISTTSTQFYKLSQELNGDMEQ 360
   |||||
Db 301 TEODLYNVISKPRNKRVPILPFVIRAGVLGRGTGISTTSTQFYKLSQELNGDMEQ 360
   |||||
QY 361 VTDSLVTLODQNSLAAVLQNRALDLLTAKGGTCLFLGEECCYVNSGIVTEKVK 420
   |||||
Db 361 VADSLVTLODQNSLAAVLQNRALDLLTAKGGTCLFLGEECCYVNSGIVTEKVK 420
   |||||
QY 421 IRDRIQRAEELQNTERWGLLSQWMPWLPFLGFLAAILLLFGPCIFNLLVKFVSSRI 480
   |||||
Db 421 IRDRIQRAEELQNTGPWGLLSQWMPWLPFLGFLAAILLLFGPCIFNLLVNFVSSRI 480
   |||||
QY 481 EAVKLQWLQME 492
   |||||
Db 481 EAVKLQWEPKMQ 492
   |||||
```

RESULT 7

US-10-016-249-4
; Sequence 4, Application US/10016249
; Publication No. US20030100053A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Mi, Sha
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 6006B.AJ172A
; CURRENT APPLICATION NUMBER: US/10/016,249
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: US/09/175,928
; PRIOR FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 62

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-016-249-4

Query Match
Best Local Similarity 88.0%; Score 2319; DB 14; Length 538;
Matches 436; Conservative 18; Mismatches 38; Indels 0; Gaps 0;

Qy 1 MALPYHTFLFTVLLPPFALTAPPCCCTTSSSPYQBFLLXRTLPGNIDAPSYBSLSKSGNS 60
Db 1 MALPYHIFLFTVLLPSFTLTAPPCCMTSSSPYQBFLLXRMQPGNIDAPSYBSLSKSGTP 60
61 TFTAHTHMPRNCYNSATLCWHANTHYWTGMINPSCPGGLGATVCWTFYFHTSMSDGGI 120
61 TFTAHTHMPRNCYNSATLCWHANTHYWTGMINPSCPGGLGATVCWTFYFHTSMSDGGV 120
121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRLTHRLVSLFNTTLRLHEVSA 180
121 QGQAREKHVKEVISQLTRVHGTSPPYKGLDLSKLHETLRLTHRLVSLFNTTLRLHEVSA 180
181 QNPTNCMCLPLHFRPYISIPVPEQNNFSTEINTTSLVGLPLVSLNLEIHTHTSNLTCVKF 240
181 QNPTNCMCLPLHFRPYISIPVPEQNNFSTEINTTSLVGLPLVSLNLEIHTHTSNLTCVKF 240
241 SNTIDTSSQICIRWVTPPTRIIVCLPSGIFVCGTSAHYCLNGSSESMCFSLFVLPMTIY 300
241 SNTYTTNSQCIRWVTPPTQIVCLPSGIFVCGTSAHYCLNGSSESMCFSLFVLPMTIY 300
301 TEODLYNVVVPKPNKRVPILPFVIRAGVLGRGTGIGSTTSTQFYKLSQBSINGDMEQ 360
301 TEODLYNVISKPRNKRVPILPFVIRAGVLGRGTGIGSTTSTQFYKLSQBSINGDMEQ 360
361 VTDLSVTLQDOLNSLAAVLQNRALDILLTAKGGTCLFLGEERCYYVQNSRIIVTEKVK 420
361 VADSLVTLQDOLNSLAAVLQNRALDILLTAKGGTCLFLGEERCYYVQNSRIIVTEKVK 420
421 IRDRIORAEELQNTBWLQNSQMPWLPFLGLPLAAILLLFGPCIFNLLVKFVSSRI 480
421 IRDRIORAEELQNTBWLQNSQMPWLPFLGLPLAAILLLFGPCIFNLLVKFVSSRI 480
481 EAVKLQWLQME 492
481 EAVKLQWEPKMQ 492

RESULT 9
US-10-133-036-21
; Sequence 21, Application US/10133036
; Publication No. US20040054133A1
; GENERAL INFORMATION:
; APPLICANT: Conrad, Bernard
; APPLICANT: Mach, Bernard
; TITLE OF INVENTION: Multiple Sclerosis-Related Supranantigen
; FILE REFERENCE: 23135-507
; CURRENT APPLICATION NUMBER: US/10/133, 036
; PRIOR FILING DATE: 2002-04-26
; PRIOR APPLICATION NUMBER: PCT/EP00/10659
; PRIOR FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Retrovirus
US-10-133-036-21

Query Match
Best Local Similarity 86.8%; Score 2285; DB 15; Length 538;
Matches 429; Conservative 21; Mismatches 42; Indels 0; Gaps 0;

Qy 1 MALPYHTFLFTVLLPPFALTAPPCCCTTSSSPYQBFLLXRTLPGNIDAPSYBSLSKSGNS 60
Db 1 MALPYHIFLFTVVPSPFTLTAPPCCMTSSSPYQBFLLXRMQPGNIDAPSYBSLSKSGTP 60
61 TFTAHTHMPRNCYNSATLCWHANTHYWTGMINPSCPGGLGATVCWTFYFHTSMSDGGI 120
61 TFTAHTHMPRNCYNSATLCWHANTHYWTGMINPSCPGGLGATVCWTFYFHTSMSDGGV 120
121 QGQAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRLTHRLVSLFNTTLRLHEVSA 180
121 QGQAREKHVKEVISQLTRVHGTSPPYKGLDLSKLHETLRLTHRLVSLFNTTLRLHEVSA 180
181 QNPTNCMCLPLHFRPYISIPVPEQNNFSTEINTTSLVGLPLVSLNLEIHTHTSNLTCVKF 240
181 QNPTNCMCLPLHFRPYISIPVPEQNNFSTEINTTSLVGLPLVSLNLEIHTHTSNLTCVKF 240
241 SNTIDTSSQICIRWVTPPTRIIVCLPSGIFVCGTSAHYCLNGSSESMCFSLFVLPMTIY 300
241 SNTYTTNSQCIRWVTPPTQIVCLPSGIFVCGTSAHYCLNGSSESMCFSLFVLPMTIY 300
301 TEODLYNVVVPKPNKRVPILPFVIRAGVLGRGTGIGSTTSTQFYKLSQBSINGDMEQ 360
301 TEODLYNVISKPRNKRVPILPFVIRAGVLGRGTGIGSTTSTQFYKLSQBSINGDMEQ 360
361 VTDLSVTLQDOLNSLAAVLQNRALDILLTAKGGTCLFLGEERCYYVQNSRIIVTEKVK 420
361 VADSLVTLQDOLNSLAAVLQNRALDILLTAKGGTCLFLGEERCYYVQNSRIIVTEKVK 420
421 IRDRIORAEELQNTBWLQNSQMPWLPFLGLPLAAILLLFGPCIFNLLVKFVSSRI 480
421 IRDRIORAEELQNTBWLQNSQMPWLPFLGLPLAAILLLFGPCIFNLLVKFVSSRI 480
481 EAVKLQWLQME 492
481 EAVKLQWEPKMQ 492

RESULT 8
US-10-133-036-2
; Sequence 2, Application US/10133036
; Publication No. US20040054133A1
; GENERAL INFORMATION:
; APPLICANT: Conrad, Bernard
; APPLICANT: Mach, Bernard
; TITLE OF INVENTION: Multiple Sclerosis-Related Supranantigen
; FILE REFERENCE: 23135-507
; CURRENT APPLICATION NUMBER: US/10/133, 036
; PRIOR FILING DATE: 2002-04-26
; PRIOR APPLICATION NUMBER: PCT/EP00/10659
; PRIOR FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Human endogenous retrovirus
US-10-133-036-2

Query Match
Best Local Similarity 87.9%; Score 2314; DB 15; Length 538;
Matches 435; Conservative 19; Mismatches 38; Indels 0; Gaps 0;
```

```
QY 181 QNPTNCWMLPLHFRPYISTPVEQWNNFSTEINTTSLVGLVPLVSNLEIHTHTSNLTCVKF 240
Db 181 QNPTNCWICLPLNFRPVISIPVEQWNNFSTEINTTSLVGLVPLVSNLEIHTHTSNLTCVKF 240
QY 241 SNTIDTSSQICIRWVTPPTIIVCLPSGIFPVCGTSAVHCLNGSSCFCFLSFLVPPMTIY 300
Db 241 SNTYTTNSQICIRWVTPPTIIVCLPSGIFPVCGTSAVHCLNGSSCFCFLSFLVPPMTIY 300
QY 301 TEQDLYNHVVPKPNKRVPIIPFVIRAGVLGRIGTGIGSTTSTQFYKLSQELNGDMEQ 360
Db 301 TEQDLYSVYISKPRNKRVPILPFVIRAGVLGALGTGIGGTTSTQFYKLSQELNGDME 360
QY 361 VTDLSVLTQDQLNSLAALVQLNRRALDILLAKRGTCCLFLGERCCYVYVNSRIVTEKVK 420
Db 361 VADSLVLTQDQLNSLAALVQLNRRALDILLAKRGTCCLFLGERCCYVYVNSRIVTEKVK 420
QY 421 IRDRIQRAEELONTERWGLSQWMPWLPFLGLPAAAILLLFGPCIFNLLVKFVSSRI 480
Db 421 IRDRIQRAEELNTPGWGLSQWMPWLPFLGLPAAAILLLFGPCIFNLLVNFVSSRI 480
QY 481 EAVKLQWLQME 492
Db 481 EAVKLQWEPKMQ 492

RESULT 10
US-10-133-036-20
; Sequence 20, Application US/10133036
; Publication No. US20040054133A1
; GENERAL INFORMATION:
; APPLICANT: Mach, Bernard
; APPLICANT: Conrad, Bernard
; TITLE OF INVENTION: Multiple Sclerosis-Related Superantigen
; FILE REFERENCE: 23135-507
; CURRENT APPLICATION NUMBER: US/10/133,036
; CURRENT FILING DATE: 2002-04-26
; PRIOR APPLICATION NUMBER: PCT/EP00/10659
; PRIOR FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 20
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Retrovirus
US-10-133-036-20

Query Match 84.9%; Score 2237; DB 15; Length 538;
Best Local Similarity 85.6%; Pred. No. 1.6e-202;
Matches 421; Conservative 26; Mismatches 45; Indels 0; Gaps 0;

QY 1 MALPYHTFLFTVLLPFPALTAPPCCCTTSSPYQEFPLXRLPGLNIDAPSYRSLSKGNS 60
Db 1 MGLPYHIFLCVLSLPCFTLTAPPCCRCWTSSPHPEFLWQRPGLNIDAPSYRSLSKGTP 60
QY 61 TFTAHTMPRNCYNSATLCMHANTHYTGKMINPSCPGGLGATVCWYTFHTSNMDSGGI 120
Db 61 TFTAHTMPRNCYNSATLCMHANTHYTGKMINPSCPGGLGATVCWYTFHTSNMDSGGV 120
QY 121 QGQAREKHVKEATSQLTRGHSTSPYKGLVLSKLHETRLHTRVLSLFTNTLTLHEVSA 180
Db 121 QGQAREKHVKEVLSQLTGTVHGTSSPYKGLDLSKLHETRLHTRVLSLFTNTLTLHEVSA 180
QY 181 QNPTNCWMLPLHFRPYISTPVEQWNNFSTEINTTSLVGLVPLVSNLEIHTHTSNLTCVKF 240
Db 181 QNPTNCWICLPLNFRPVISIPVEQWNNFSTEINTTSLVGLVPLVSNLEIHTHTSNLTCVKF 240
QY 241 SNTIDTSSQICIRWVTPPTIIVCLPSGIFPVCGTSAVHCLNGSSCFCFLSFLVPPMTIY 300
Db 241 SNTYTTNSQICIRWVTPPTIIVCLPSGIFPVCGTSAVHCLNGSSCFCFLSFLVPPMTIY 300
QY 301 TEQDLYNHVVPKPNKRVPIIPFVIRAGVLGRIGTGIGSTTSTQFYKLSQELNGDMEQ 360
Db 301 TEQDLYSVYISKPRNKRVPILPFVIRAGVLGALGTGIGGTTSTQFYKLSQELNGDME 360
```

```
Db 301 TEQDLYSVYISKPRNKRVPILPFVIRAGVLGALGTGIGGTTSTQFYKLSQELNGDME 360
QY 361 VTDLSVLTQDQLNSLAALVQLNRRALDILLAKRGTCCLFLGERCCYVYVNSRIVTEKVK 420
Db 361 VADSLVLTQDQLNSLAALVQLNRRALDILLAKRGTCCLFLGERCCYVYVNSRIVTEKVK 420
QY 421 IRDRIQRAEELONTERWGLSQWMPWLPFLGLPAAAILLLFGPCIFNLLVKFVSSRI 480
Db 421 IPRIQRIAEELNTPGWGLSRMPWLPFLGLPAAAILLLFGPCIFDILLVNFVSSRI 480
QY 481 EAVKLQWLQME 492
Db 481 EAVKLQWEPKMQ 492

RESULT 11
US-09-864-761-37656
; Sequence 37656, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37656
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002346.1
```

```
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
; OTHER INFORMATION: EST HUMAN HIT: AUI38405.1, EVALUE 7.00e-90
; OTHER INFORMATION: SWISSPROT HIT: P10269, EVALUE 1.00e-16
US-09-864-761-37656

Query Match      46.9%; Score 1235.5; DB 9; Length 263;
Best Local Similarity 89.4%; Pred. No. 4e-108;
Matches 236; Conservative 8; Mismatches 19; Indels 1; Gaps 1;

QY 126 EKQVKEAISLTGRHSTPSYKGLVLSKHLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTN 185
DB 1 EKHIKEVISQLTWHSSTPGYKGLDLSKHLHETLHHTGLVSLFNTTLTGLHEVSAQNPTN 60

QY 186 CWMCLPLHFRPYISIPVEQWNNFSTEINTTSLVGLPLVSNLEITHTSNLTVCVKFSNTID 245
DB 61 CWMCLPLHFRPYISIPVEQWNNFSTEINTTSLVGLPL-SNLEITHTSNLTVCVKFSNTID 119

QY 246 TTSSQCIIRWTPTRIVCLPSGIFVCGTSAYHCLNGSSSFCFLSFLVPPMTIYTEQDL 305
DB 120 TANSQCIIRWTPTRIVCLPSGIFVCGTSAYHCLNGSSSFCFLSFLVAPMPIYTEQDL 179

QY 306 YNHVVKPHNKRVPILPFVIRAGVLGRGTGIGSITTSQFYKLSQEIINGDMQVTDLSL 365
DB 180 YNHVVKPHNKRVPILPFVIRAGVLGRGTGIGSITTSQFYKLSQEIINGDMQVTDLSL 239

QY 366 VTLQDQLNSLAAYVLQNRALDILL 389
DB 240 VTLQDQLNSLAAYVLQNRALDILL 263

RESULT 12
US-10-363-616-473
; Sequence 473, Application US/10363616
; Publication No. US20040044181A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-113 (793)
; CURRENT APPLICATION NUMBER: US/10/363,616
; PRIOR FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: 09/654,935
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 490
; SEQ ID NO 473
; LENGTH: 531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-363-616-473

Query Match      42.6%; Score 1122; DB 15; Length 531;
Best Local Similarity 51.6%; Pred. No. 6.2e-97;
Matches 246; Conservative 6; Mismatches 29; Indels 196; Gaps 5;

QY 36 EFLKRTPLGNIAPSYRSLSKGNSTPTAHTHPRN-----C-----YNSATLCMH- 81
DB 112 EFGNKKQRNLQATSLISCAT-----VLSTRSPRCWQVCLVKAILFFQDDTLHLHP 166

QY 82 -----ANTHYWTGKMINPSCFGLGATVCWTFYFHTSMDSGGIGQGAKEKQVKEAISQL 136
DB 167 PEGNAAPFYGGGKMINPSCFGLGATVCWTFYFHTSMDSGGIGQ----- 211

QY 137 TRGHSTSPYKGLVLSKHLHETLRTHTRLVSLFNTTLTRLHEVSAQNPTNCWMLPLHFRP 196
DB 212 ----- 211
```

```
QY 197 YISIPVEQWNNFSTEINTTSLVGLPLVSNLEITHTSNLTVCVKFSNTIDTTSSQCIIRWVT 256
DB 212 ----- 211

QY 257 PPTRIVCLPSGIFVCGTSAYHCLNGSSSFCFLSFLVPPMTIYTEQDLYNHVVPKPHNK 316
DB 212 -----VPMTIYTEQDLYNHVVPKPHNK 234

QY 317 RVPILPFVIRAGVLGRGTGIGSITTSQFYKLSQEIINGDMQVTDLSLTLQDQLNSLA 376
DB 235 RVPILPFVIRAGVLGRGTGIGSITTSQFYKLSQEIINGDMQVTDLSLTLQDQLNSLA 294

QY 377 AVVLQNRALDILLTAKRGGTCLFLGBERCCYVYVNSQSRIVTEKVKIIRDRIOCRABEELONTE 436
DB 295 AVVLQNRALDILLTAKRGGTCLFLGBERCCYVYVNSQSRIVTEKVKIIRDRIOCRABEELONTE 354

QY 437 RWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFENLLVKKFVSSRIEAVKQLQWLOMEP 493
DB 355 HWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFENLLVKKFVSSRIEAVKQLQWLOMEP 411

RESULT 13
US-10-133-036-4
; Sequence 4, Application US/10133036
; Publication No. US20040054133A1
; GENERAL INFORMATION:
; APPLICANT: Mach, Bernard
; TITLE OF INVENTION: Multiple Sclerosis-Related Superantigen
; FILE REFERENCE: 23135-507
; CURRENT APPLICATION NUMBER: US/10/133,036
; PRIOR FILING DATE: 2002-04-26
; PRIOR APPLICATION NUMBER: PCT/EP00/10659
; PRIOR FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Human endogenous retrovirus
US-10-133-036-4

Query Match      22.4%; Score 591; DB 15; Length 120;
Best Local Similarity 86.7%; Pred. No. 1.4e-47;
Matches 104; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 1 MALPYHTFLFTVLLPPFALTAPPCCCTTSSSSPYQBFELXETRLPGNIDAPSYLSKSGNS 60
DB 1 MALPYHTFLFTVLLPSFTLTAPPCCRCMTSSSSPYQBFELXETRLPGNIDAPSYLSKSGTP 60

QY 61 TFTAHTHMPRNCYNSATLCHMANTHYWTGKMINPSCFGLGATVCWTFYFHTSMDSGGGI 120
DB 61 TFTAHTHMPRNCYNSATLCHMANTHYWTGKMINPSCFGLGATVCWTFYFHTSMDSGGGV 120

RESULT 14
US-08-979-847-109
; Sequence 109, Application US/08979847
; Publication No. US20030039664A1
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
; APPLICANT: TUKE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
```


US-10-001-835-168

Query Match 19.5%; Score 514.5; DB 14; Length 576;
Best Local Similarity 25.5%; Pred. No. 2.4e-39;
Matches 142; Conservative 63; Mismatches 123; Indels 229; Gaps 16;

QY 140 HSTPSPYKGLVLSKLHETLR--THRLVSLFNTLTLRLHEVSAONPT---NCWMLPLHFR 195
Db 26 HPLPLP-----SYLHHTINLTHSL--AASNPLVNNWMLCISLSS 65
QY 196 PYISIP-VPBQW----- 206
Db 66 AYTAFAVQTDWATSPISLHLRTSPNSPHLYPPEELIYFLDRSSKTSPIDISHQAAALLR 125
QY 207 ---NPFSTEINTSVLVGPLVS----- 225
Db 126 TYLKNLSPYINSTPIFGPLTOTTIPVAAPLCISWQRPCTGIPGLNLSFRCSTLHLRS 185
QY 226 -----NLEITHSNLTCVKSNTIDTSSQCIR----- 253
Db 186 PTTNINETIGAFQLHITDKFSINTDKLN---ISSNYCLGRHLPCLSLHPWLSPPCSSDS 242
QY 254 ----- 253
Db 243 PPRPSSCLLIPSPENNSERLLVDRFLIHENRTTPSTQLPHOSPLQPLTAALAGSLG 302
QY 254 -WV---TP---PTRI-----VCLPSGIFVCGTSAYHCLNGSSECMFSLFVPPMTIY- 300
Db 303 VWQDTPFSPSHLFTLHLQFCLAQGLFLCGSSTYMCLPANWTGCTLVFLTPKIQFAN 362
QY 301 -TEQDLYNHVVPKPKVPIVIRAGVLG---RLGTGIGSITTSQFYKLSOEING 356
Db 363 GTELPVPLMTPTQQRKVIPLIPLMVGLGUSASTVALGTGAGISTVWTFRSLNDFS 422
QY 357 DMEQVDSLVTDQNLNSLAAVLQNRRLDLTAKRGGTCLFLGBCRYVYNQSRIVTE 416
Db 423 SITDISQTLVLQAVDLSLAAVLQNRERGLDLTAEGKGLCIFNEECFYLQSGLVYD 482
QY 417 KVKEIRDRIQRAEELQNTER--WGLLSQWMPVLPFLGPIALILLLFGPCIFNLLVK 474
Db 483 NIKKLDRAQKLANQASNAEPPWA-LSNWSMWLPVTSPLIFLILLFGPCIFRLVSQ 541
QY 475 FVSSRIEAVKLQVLOM 491
Db 542 FIQNRIOAITNHSIRQM 558

RESULT 19
US-10-001-835-168
; Sequence 168, Application US/10001835
; Publication No. US20020160387A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Caferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro
; FILE REFERENCE: DEX-0277
; CURRENT APPLICATION NUMBER: US/10/001,835
; PRIOR FILING DATE: 2001-11-20
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 168
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Homo sapien

US-10-001-835-168

Query Match 18.2%; Score 480.5; DB 13; Length 469;
Best Local Similarity 31.4%; Pred. No. 3e-36;
Matches 132; Conservative 62; Mismatches 141; Indels 85; Gaps 15;

QY 136 LTRGHSTPSPYKGLVLSKLHETLR-----TH-TRLVSLFNTLT-----RLHEVSAQ 181
Db 53 ISQRPTGIPGLNLSFRCSTLHLRSPTTHITETIGAFQLHITDKFSINTDKLKNISSN 112
QY 182 NPTNCWMLPLHF-----RPYISIPVPEQWNNFSTEI----- 213
Db 113 -----YCLGRHLPLSILHPWLPSPCSSDSPRPSSRLILPSPKNSERLLVDTQRFILH 166
QY 214 --NTTSVLVGPLVSNLEITHTS---NLTCVKFSNTIDTSSQCIRWV-----TPPTRIV 262
Db 167 HENRTS-----PSTQLPHQSPLOPLTAASLAGSLG1-----WVQDTPFSTPHLFTL 212
QY 263 ----CLPSGIFVCGTSAYHCLNGSSECMFSLFVPPMTIY---TEQDLYNHVVPKPHNK 316
Db 213 HLQFCLTQGLFLCGSSTYMCLPANWTGCTLVFLTPKIQFANGTEELPVPLMTPTQKR 272
QY 317 RVPILPFVIRAGVLG---RLGTGIGSITTSQFYKLSOEINGDMEQVDSLVTDQQLN 373
Db 273 VIPLIPLMVGLGUSASTIALGTGAGISTVTFRSLNDFSASITDISQTLVLQAVD 332
QY 374 SLAAVVLQNRRLDLTAKRGGTCLFLGBCRYVYNQSRIVTEKVKEIRDRIQRAEELQ 433
Db 333 SLAAVVLQNRRLDLTAKRGGTCLFLGBCRYVYNQSRIVTEKVKEIRDRIQRAEELQ 392
QY 434 NTER--WGLLSQWMPVLPFLGPIALILLLFGPCIFNLLVKFVSRIEAVKLQVLOM 491
Db 393 NYAEPPWA-LSNWSMWLPVTSPLIFLILLFGPCIFRLVSQFIQNRIOAITNHSIAQM 451

RESULT 20
US-08-979-847-81
; Sequence 81, Application US/08979847
; Publication No. US20030039664A1
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
; APPLICANT: TUKE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTI
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,847
; FILING DATE: 26-NOV-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BERRIDGE, WILLIAM P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 39046A

; Sequence 5, Application US/09954983
; Publication No. US20020137889A1
; GENERAL INFORMATION:
; APPLICANT: SOONG, NAY WEI
; APPLICANT: STEMMER, WILLEM P.
; APPLICANT: POWELL, SHARON K.
; APPLICANT: OTTO, EDWARD
; TITLE OF INVENTION: STRESS RESISTANT RETROVIRUSES
; FILE REFERENCE: 02-107120PC
; CURRENT APPLICATION NUMBER: US/09/954,983
; CURRENT FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/233,398
; PRIOR FILING DATE: 2000-09-18
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Clone ID 4-4
US-09-954-983-5

Query Match 14.7%; Score 386; DB 9; Length 672;
Best Local Similarity 23.9%; Pred. No. 4.5e-27;
Matches 144; Conservative 76; Mismatches 219; Indels 164; Gaps 22;
QY 16 PFALTAPPCCCTSSS-----PYQEFKXRLTLPON---IDAPSVRSLSKGNSTF 62
DB 94 PYSPPGPPCCSGSSGNSACFRDCDEPLTSLPRNTAWNRKLQDVTHKS-SEGIFYVC 152
QY 63 TATHMPR---NC-----YNSATLCMHANTHYWTGMINPSCPGGLGATVCTWTFHTS 113
DB 153 PG-PHREAKSCGGPDSFYCASGCGCTTGRVW-----KPS-----SMDYIT-- 195
QY 114 MSGGGIGQQA-----REKVKRAISQLTRGH----- 140
DB 196 -VDNLTNTQAVKVKDKNCNPLAIRFTNTGRQVTSWTTGHWGLRLYVTGDKPGLTFG 254
QY 141 -----STSPYKGLVLSKLHETL----- 158
DB 255 IRLKYQNLGRVPIGNPNVLADQLSPFLPNLPKPAKSPASNSTPTLISPSPTPTQPPP 314
QY 159 -RTHRLVSLFNTTLRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQMNFSSTEINTTS 217
DB 315 AGTGDRLNLVQAYQALNTPDKTQECWCLVSGPPYEGVAV-----LGTYSNHTS 368
QY 218 VLVGPLVSNLEITHTSNLTVCKFSNTIDTSSQCIKRWTPPTTRIVC----- 263
DB 369 ---APANCSTASOKHLTSEV-----TGRGLCIGTVPKTHQALCNTLLTKGKSYLV 418
QY 264 LPSGIFVCGTSAHYCLNGS-----SESMCLFLSLVPPMTIYTEQDLNVHVVPKHNKRP 319
DB 419 APAGTMMACNTGLTPCLSATVLNRTDYCVLVELWPRTVTHPPSYVYSQFENSRYRHKREP 478
QY 320 I---LPFVIRAGVLGRGTGIGSITT---STQFYKLSQBIINGMEQVTDLSLVTLPQDLN 373
DB 479 VSLTALLLGLTWGGIAGVGTTALVATQFQQLHAQVODDLKEVEKSIINLEKSLT 538
QY 374 SLAAVQLQNRALDLTAKGGTCLFLGEBRCYVYNQSRIVTEKVKIIRDIQCRABELQ 433
DB 539 SLSEVVLQNRGLDLFLKEGGLCAALKECCFYADHTGLVRDSMAKRLRLTQROKLF 598
QY 434 NTERW--GLISQMPW-----VLPLGLPALAILLLFPGCIFNLLVKFVSRIEAVKLOM 487
DB 599 SSQGFELFNR-SPWFTTLLISTMGPLIILLLLFPGCILRLRLVQFVKDRISVVOALV 657
QY 488 VLQ 490
DB 658 LTQ 660

RESULT 30
US-09-954-983-7
; Sequence 7, Application US/09954983
; Publication No. US20020137889A1
; GENERAL INFORMATION:
; APPLICANT: SOONG, NAY WEI
; APPLICANT: STEMMER, WILLEM P.
; APPLICANT: POWELL, SHARON K.
; APPLICANT: OTTO, EDWARD
; TITLE OF INVENTION: STRESS RESISTANT RETROVIRUSES
; FILE REFERENCE: 02-107120PC
; CURRENT APPLICATION NUMBER: US/09/954,983
; CURRENT FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/233,398
; PRIOR FILING DATE: 2000-09-18
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 675
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Clone ID 2B-13
US-09-954-983-7

Query Match 14.7%; Score 386; DB 9; Length 675;
Best Local Similarity 23.9%; Pred. No. 4.5e-27;
Matches 144; Conservative 76; Mismatches 219; Indels 164; Gaps 22;
QY 16 PFALTAPPCCCTSSS-----PYQEFKXRLTLPON---IDAPSVRSLSKGNSTF 62
DB 97 PYSPPGPPCCSGSSGNSACFRDCDEPLTSLPRNTAWNRKLQDVTHKS-SEGIFYVC 155
QY 63 TATHMPR---NC-----YNSATLCMHANTHYWTGMINPSCPGGLGATVCTWTFHTS 113
DB 156 PG-PHREAKSCGGPDSFYCASGCGCTTGRVW-----EPSS-----SMDYIT-- 198
QY 114 MSGGGIGQQA-----REKVKRAISQLTRGH----- 140
DB 199 -VDNLTNTQAVKVKDKNCNPLAIRFTNTGRQVTSWTTGHWGLRLYVTGDKPGLTFG 257
QY 141 -----STSPYKGLVLSKLHETL----- 158
DB 258 IRLKYQNLGRVPIGNPNVLADQLSPFLPNLPKPAKSPASNSTPTLISPSPTPTQPPP 317
QY 159 -RTHRLVSLFNTTLRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQMNFSSTEINTTS 217
DB 318 AGTGDRLNLVQAYQALNTPDKTQECWCLVSGPPYEGVAV-----LGTYSNHTS 371
QY 218 VLVGPLVSNLEITHTSNLTVCKFSNTIDTSSQCIKRWTPPTTRIVC----- 263
DB 372 ---APANCSTASOKHLTSEV-----TGRGLCIGTVPKTHQALCNTLLTKGKSYLV 421
QY 264 LPSGIFVCGTSAHYCLNGS-----SESMCLFLSLVPPMTIYTEQDLNVHVVPKHNKRP 319
DB 422 APAGTMMACNTGLTPCLSATVLNRTDYCVLVELWPRTVTHPPSYVYSQFENSRYRHKREP 481
QY 320 I---LPFVIRAGVLGRGTGIGSITT---STQFYKLSQBIINGMEQVTDLSLVTLPQDLN 373
DB 482 VSLTALLLGLTWGGIAGVGTTALVATQFQQLHAQVODDLKEVEKSIINLEKSLT 541
QY 374 SLAAVQLQNRALDLTAKGGTCLFLGEBRCYVYNQSRIVTEKVKIIRDIQCRABELQ 433
DB 542 SLSEVVLQNRGLDLFLKEGGLCAALKECCFYADHTGLVRDSMAKRLRLTQROKLF 601
QY 434 NTERW--GLISQMPW-----VLPLGLPALAILLLFPGCIFNLLVKFVSRIEAVKLOM 487
DB 602 SSQGFELFNR-SPWFTTLLISTMGPLIILLLLFPGCILRLRLVQFVKDRISVVOALV 660
QY 488 VLQ 490
DB 661 LTQ 663

Search completed: January 22, 2005, 22:17:05
Job time : 110 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2005, 21:17:11 ; Search time 39 Seconds
(without alignments)
1216.279 Million cell updates/sec

Title: US-09-319-156B-10
Perfect score: 2634
Sequence: 1 MALPYHTFLTVLLPPFALT.....KFVSSRIEAVKLMQVLMQEP 493
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	580	22.0	563	1 VCMWV7	env polyprotein -
2	559.5	21.2	567	1 VCFVAS	env polyprotein -
3	556	21.1	587	1 VCLJSA	env polyprotein -
4	546	20.7	586	1 VCLJMP	env polyprotein -
5	477.5	18.1	575	1 VCLJHD	env polyprotein pr
6	476.5	18.1	570	4 B44282	retrovirus-related
7	463.5	17.6	582	1 VCVBAR	env polyprotein -
8	415.5	15.8	321	2 S12812	env polyprotein -
9	409	15.5	676	2 T01381	env protein - muri
10	407.5	15.5	665	1 VCVNEM	env polyprotein -
11	402.5	15.3	636	1 VCMVFS	env polyprotein -
12	400	15.2	688	2 A43491	env polyprotein -
13	398	15.1	689	2 B43491	env polyprotein -
14	393	14.9	676	1 VCMVPV	env polyprotein pr
15	392.5	14.9	661	1 VCMVCB	env polyprotein -
16	392	14.9	676	2 S70395	env polyprotein -
17	391.5	14.9	662	1 VCMVGF	env polyprotein -
18	391.5	14.9	662	2 A25982	env polyprotein -
19	390.5	14.8	640	1 VCMVRV	env polyprotein pr
20	390	14.8	669	1 VCVWEK	env polyprotein -
21	389.5	14.8	662	1 VCMVLB	env polyprotein -
22	386	14.7	627	1 VCMVM2	env polyprotein -
23	385	14.6	668	2 VCMVFP	env polyprotein pr
24	385	14.6	669	2 A46511	envelope protein -
25	384	14.6	640	1 VCMVM1	env polyprotein -
26	383	14.5	642	1 VCMVS2	env polyprotein pr
27	383	14.5	645	1 VCMVSS	env polyprotein pr
28	382	14.5	642	2 T10533	env polyprotein pr
29	380.5	14.4	639	1 VCMVSA	env polyprotein pr

RESULT 1
VCMWV7
env polyprotein - baboon endogenous virus (strain M7)
N;Alternate names: coat polyprotein
C;Contains: coat protein gp70; coat protein p20E
C;Species: baboon endogenous virus
A;Note: host Papio sp. (baboon)
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C;Accession: JT0262
R;Kato, S.; Matsuo, K.; Nishimura, N.; Takahashi, N.; Takano, T.
Jpn. J. Genet. 62, 127-137, 1987
A;Title: The entire nucleotide sequence of baboon endogenous virus DNA: a chimeric genome
A;Reference number: JT0260
A;Accession: JT0262
A;Molecule type: DNA
A;Residues: 1-563 <KAT>
A;Cross-references: UNIPROT:P10269; GB:M16550; NID:G509586; PIDN:AAA87333.1; PID:G332599
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; transmembrane protein
F;1-376/Product: coat protein gp70 #status predicted <CGP>
F;377-563/Product: coat protein p20E #status predicted <CGP>
F;437-449/Region: immunosuppressive peptide #status predicted
F;113,219,229,264,282,292,306,312,321,339,469/Binding site: carbohydrate (Asn) (covalent)

ALIGNMENTS

30	380.5	14.4	666	1 VCMVHL	env polyprotein pr
31	379	14.4	665	1 VCMVVR	env polyprotein pr
32	377	14.3	665	1 VCMVKA	env polyprotein pr
33	376	14.3	642	1 VCMVFG	env polyprotein -
34	367	13.9	644	2 S15464	gp70 protein - mur
35	364.5	13.8	353	2 S22805	env polyprotein -
36	359	13.6	671	1 VCMVCE	env polyprotein -
37	357.5	13.6	667	1 VCLJGL	env polyprotein pr
38	311.5	11.8	201	2 A03985	env polyprotein -
39	311.5	11.8	211	2 A03986	env polyprotein -
40	226	8.6	488	2 A45714	envelope glycoprot
41	217	8.2	488	1 VCLJCN	env polyprotein -
42	209	7.9	488	1 VCLJMT	env polyprotein -
43	208	7.9	488	2 S14605	envelope glycoprot
44	204	7.7	486	1 VCLJH2	env polyprotein -
45	203	7.7	488	1 VCMVH	env polyprotein -

```
Db 352 DIDIIPGDEVPPIAIDHFIYRP-KRAIQFIPLLAGIGITAAFTTGATGLGVSVTQYTKL 410
QY 351 SOBINGMEQVDSLVTLQDLNSLAADVQLNRRALDLLTAKRGGTCFLGGERCYVYVNO 410
Db 411 SNQLISDVQLSTIQDLQDVDSLAEVVLQNRGLDLLTAEQGGICLALQKCCFYVVK 470
QY 411 SRIVTEKVEIRDRIQCRABELQNTERWGLLSQWMPVLPFLGPLAALILLLLFGPCIFN 470
Db 471 SGIVRDIKTQBELERRRDLASNPWTGLQGLLPYLLPFLGPLTLLLLLTIGPCIFN 530
QY 471 LLVKFVSSRIEAVKLQWVLQ 490
Db 531 RLTAFINDLNLIITHAMVLQ 550

RESULT 2
VCFVAS
env polyprotein - avian spleen necrosis virus
N;Alternate names: coat polyprotein
N;Contains: coat protein gp22; coat protein gp73
C;Species: avian spleen necrosis virus
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 09-Jul-2004
C;Accession: A38212
R;Kewalramani, V.N.; Panganiban, A.T.; Emerman, M.
J. Virol. 66, 3026-3031, 1992
A;Title: Spleen necrosis virus, an avian immunosuppressive retrovirus, shares a receptor
A;Reference number: A38212; MUID:92219390; PMID:1313915
A;Accession: A38212
A;Molecule type: DNA
A;Residues: 1-567 <NEW>
A;Cross-references: UNIPROT:P31796; GB:M87666
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; transmembrane protein
F;1-401/Domain: extracellular #status predicted <EXT>
F;1-397/Product: coat protein gp73 #status predicted <CP1>
F;363-379/Region: hydrophobic #status predicted
F;394-397/Region: cleavage processing #status predicted
F;398-567/Product: coat protein gp22 #status predicted <CP2>
F;402-418/Domain: transmembrane #status predicted <TM1>
F;419-567/Domain: intracellular #status predicted <INT>
F;245,274,306,328,335/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 21.2%; Score 559.5; DB 1; Length 567;
Best Local Similarity 31.9%; Pred. No. 5.5e-35;
Matches 153; Conservative 75; Mismatches 158; Indels 93; Gaps 19;

QY 66 THMPCYNYSATLCMHANTHYWT-----GKMINPSCPGGLGATVCWITYFT 110
Db 121 TOMHSTCYEKTQECTLLGKTYFTAILQKTKLGSYEDGNPKLIQASCTGTGKPCVCDPVA 180
QY 111 HTSMDSGGGLQGAQREKQVKEAISQLTRGHSTPS-PYKGLVLSKLHETLTHRLVSLFN 169
Db 181 PVYVSDGGGTDMIRESVREERLEIR-HSYSPVQHPHALPR-----SRGVDLDP 231
QY 170 TT---LTRLHEV-SAQNP---TNCWMLPLHFRPYISIPVPEQW-----NFTSEI--- 213
Db 232 QTSDLLEATHQVLNATNPKLAENCWLMTL--GTPPAIPTNGVTLNGLNCSLSLPFGC 289
QY 214 -----NTTSVLVGPLVSNLEITHNLNLCVKSNTDITTSQCIRWVTP 257
Db 290 NPPGSDVSCYAGEADNRGTGIPVG-----YVHFTNCTSIQ---EVTNETSQ---MGN 335
QY 258 PTRIVCLPSGIFVCGTSS-AYHCLNGSSESMCLFLPVPMTIYTQDLNHNHVPKP--- 313
Db 336 LTRL-CPPPGHVFVCGNNMAYTALPNKWIGLCILASIVPDISIISGEE---PIPLPSIE 390
QY 314 ----HNKRPILPFVIRAGVGLR---GTGIGSIITTSFYKLSQEIINGDMEQVTDLS 365
Db 391 YTARRHKRAVQFIPLLVGLGISATLAGGTGLG---VSVHTYHKLNSQLIEDVQALSGTI 447
QY 366 VTLDQDLNSLAADVQLNRRALDLLTAKRGGTCFLGGERCYVYVNOQRIVTEKVEIRDRI 425
```

```
Db 448 NDLDQDIDSIAEVLQNRREGLDLLTAEQGGICLALQKCCFYANKSGIVRDKIRKQEDL 507
QY 426 QCRABELQNTERWGLLSQWMPVLPFLGPLAALILLLLFGPCIFNLLKVPSSRIEAVK 484
Db 508 IERKRALYDNPLWGLNGFLPYLLPFLGPLFGLILFELTGPCIMKMTLTRIHKIQAVK 566

RESULT 3
VCLTSA
env polyprotein - simian AIDS retrovirus SRV-1
N;Alternate names: coat polyprotein
N;Contains: coat protein gp20; coat protein gp70
C;Species: simian AIDS retrovirus SRV-1
A;Note: host Macaca mulatta (rhesus macaque)
C;Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 16-Jul-1999
C;Accession: A04000
R;Power, M.D.; Marx, P.A.; Bryant, M.L.; Gardner, M.B.; Barr, P.J.; Luciw, P.A.
Science 231, 1567-1572, 1986
A;Title: Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrom
A;Reference number: A94711; MUID:86151668; PMID:3006247
A;Accession: A04000
A;Molecule type: DNA
A;Residues: 1-587 <POW>
A;Cross-references: GB:M11841; NID:G334746; PIDN:AAA47733.1; PID:G334750
C;Comment: The env polyprotein contains coat protein gp70 and coat protein gp20; however,
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: capsid protein; coat protein; glycoprotein; polyprotein
F;120,237,266,271,277,280,295,308,322,328,340,358,488,585/Binding site: carbohydrate (Asn)

Query Match 21.1%; Score 556; DB 1; Length 587;
Best Local Similarity 29.4%; Pred. No. 1.1e-34;
Matches 160; Conservative 84; Mismatches 201; Indels 99; Gaps 17;

QY 19 LTAPPPCCCTT-SSSPYQBFLLXTRL-----PGNIDAPYSRLSKG 58
Db 49 VSSPPTNLTTVTSCSYTAYSVTNSLKWQCVSPPTTASPTHIGSCPCSNQSYDSVH-- 106
QY 59 NSTFTTAHTMPRCNYSATLCMHANTHYWTGKI---NPS----- 95
Db 107 -----ATCYNHYOQCTIGNKTYTATMIRKSPSSGSGDNVPTILGNQNLIAG 155
QY 96 CP-GGLGATVCWITYFTHTSMDSGGGLQGAQREKQVKEAISQLTRGHSTPSPYKGLVLSKL 154
Db 156 CPENKKGQVVCWNSQSPSVHSDGGGPDQKVREIIVNKKFEELHKLFPPELSYHPLALPEA 215
QY 155 H--ETLRTHT-----RLVSLFNLTTLTRLHEVSAQNPNTNCWMLPLHFRPYISIPVPEQW 207
Db 216 RGKEKIDAHTFDLLATVHSLNVSQR-----QLAEDCWLCI-----RSGDPVPLALPYD 265
QY 208 NESTEINT-----TSVLGP--LVSNLEITHNTSLNLCVKFSNTID-----TTSQ 250
Db 266 NTSCSNSTFFNCSNCSCLITPFLVQPNFTHSVCLYADYQNNSFDDIVGLAGFTNCSS 325
QY 251 CIRWVTPTRIVCLPSGIFVCGTSS-AYHCLNGSSESMCLFLPVPMTIYTQD----- 304
Db 326 YIN-ISKSPSLCAPNSVFCVGNKAYTLPNWTGSCVLAALLDIDIPGSEVPVP 384
QY 305 LYNHVVPKPNKRVPLPVPVIRAGVGLRGTGIGSTTTSTQFYKLSQEIINGDMEQVTD 364
Db 385 AIDHFLGRP-KRAIQFIPLVIGLITVAVSTGTAGLGVSLTQYTKLSHQLISDVQAISST 443
QY 365 LVTLQDLNSLAADVQLNRRALDLLTAKRGGTCFLGGERCYVYVNOQRIVTEKVEIRD 424
Db 444 IODLQDVDSLAEVVLQNRREGLDLLTAEQGGICLALQKCCFYANKSGIVRDKIRKQEDD 503
QY 425 IQCRABELQNTERWGLLSQWMPVLPFLGPLAALILLLLFGPCIFNLLKVPSSRIEAVK 484
Db 504 LEKRRQLIDNPFTWTFGHGLLPVYVMPLLGPLLCLLLVLSFGPIIFNKLMTFIKHQIESIQ 563
QY 485 LQMV 488
```

Db 564 AKPI 567

RESULT 4
VCLJMP
env polyprotein - Mason-Pfizer monkey virus
N;Alternate names: coat polyprotein
N;Contains: coat protein gp20; coat protein gp70
C;Species: Mason-Pfizer monkey virus
C;Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 09-Jul-2004
C;Accession: D25839
R;Sonigo, P.; Barker, C.; Hunter, E.; Wain-Hobson, S.
Cell 45, 375-385, 1986
A;Title: Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type virus
A;Reference number: A90878; MUID:86189951; PMID:2421920
A;Accession: D25839
A;Molecule type: DNA
A;Residues: 1-586 <SON>
A;Cross-references: UNIPROT:P07575; GB:M12349; NID:G334702; PIDN:AAA47712.1; PID:G334705
A;Experimental source: clone 6A
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: capsid protein; coat protein; glycoprotein; polyprotein
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-394/Product: coat protein gp70 #status predicted <GP7>
F;395-586/Product: coat protein gp20 #status predicted <GP2>
F;120,237,264,276,291,304,318,324,339,357,487/Binding site: carbohydrate (Asn) (covalent)

Query Match 20.7%; Score 546; DB 1; Length 586;
Best Local Similarity 28.6%; Pred. No. 6.2e-34;
Matches 145; Conservative 91; Mismatches 209; Indels 62; Gaps 13;

QY 26 CCTSSSPYQEFKXRLPGNIDAPSVRSKGNSTFTATHMPNRCYNATLCMHANTH 85
Db 78 CVSTPTPSNTHG--SCPECTNYSVSH-----ASCNHYQCCNTGNKT 122

QY 86 YWTG-----KMINPSCPG-LGATVCWTFYTHSTMSDGGGIQ 121
Db 123 YLTATITGDRTPAIGDGNVPTVLGTSNHLITAGCPNGKGQGVCMNRPSPVHSDGGPQ 182

QY 122 GQAREQVKEAISQLRGHTSPYKGLVLSKLH--ETLTHRLVSLFNTLTRHEVS 179
Db 183 DKARDIIVNKKFELHRSFPFELSYPHALPEARKEKIDAHF--LDLLATVHSLNASQ 240

QY 180 AQNPNCWMLPLHFRPYISIPVPEQ--WNNFSTEINTSVLVGP-LVSNLEITHTSNLT 237
Db 241 PSIAEDCNLCQSGDVPVLPALPNDTLCSNFCALSNHSCPLTPPFLVQPPNFTDNCLYA 300

QY 238 VKFSNTID-----TTSQC-----IRWVTPTRIVCLPSGIFVCGTS--AYHCLNGSSES 286
Db 301 HYQNNSFDIDVGLASFTNCSSYVNVSTASPSNSLCPNSVFCVCGNNKAYTYLPTNWTG 360

QY 287 MCFSLFVPPMTIYTED-----LYNHVPKPKNKRPVLPFVIRAGVLRGLTGTSIT 341
Db 361 SCVLATLPLPDIDIPGSEPVPIPAIDHFLGKA--KRAIQLPLFVGLGITTAFTAVTGAAGLG 419

QY 342 TSTQFYKLSQIEINGMEQVTDLSVTLQDOLNSLAAVLQNRALDILTAKEGTCGLFLG 401
Db 420 VSTQYTKLHQISDVQALISSTIQDLQDQVDSLAEVLLQNRGLDILTAEOQGICLALQ 479

QY 402 EERCYVYVNSRIVTEKVEIRDIQRAEQLQTERWGLLSQWMPVLPFLGFLAALIL 461
Db 480 EKCCFFVANKSGIVRDKIKNLQDLERRRRQLINDPFTWSPHGFPLPYWMLPLGLLCLLV 539

QY 462 LLFGPCIFNLVVKVSSRIEAVKLQWV 488
Db 540 LSPGPIIFNKLMTFIKHQIESIQAKPI 566

RESULT 5
VCLJHD

env polyprotein precursor - squirrel monkey retrovirus SMRV-H
N;Alternate names: coat polyprotein
N;Contains: coat protein gp20; outer membrane protein
C;Species: squirrel monkey retrovirus SMRV-H
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 09-Jul-2004
C;Accession: D31827
R;Oda, T.; Ikeda, S.; Watanabe, S.; Hatushika, M.; Akiyama, K.; Mitsunobu, F.
Virology 167, 468-476, 1988
A;Title: Molecular cloning, complete nucleotide sequence, and gene structure of the provirus
A;Reference number: A31827; MUID:89073750; PMID:3201749
A;Accession: D31827
A;Molecule type: DNA
A;Residues: 1-575 <ODA>
A;Cross-references: UNIPROT:P21412; GB:M23385; NID:G332626; PIDN:AAA66455.1; PID:G332627
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: capsid protein; coat protein; glycoprotein; polyprotein; transmembrane protein
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-386/Product: outer membrane protein #status predicted <OMP>
F;387-575/Product: coat protein gp20 #status predicted <TMP>
F;387-403/Domain: transmembrane #status predicted <TM1>
F;447-481/Region: immunosuppressive peptide
F;518-534/Domain: transmembrane #status predicted <TM2>
F;126,239,266,271,302,316,322,349,479/Binding site: carbohydrate (Asn) (covalent) #status

Query Match 18.1%; Score 477.5; DB 1; Length 575;
Best Local Similarity 28.5%; Pred. No. 1.1e-28;
Matches 136; Conservative 81; Mismatches 168; Indels 93; Gaps 16;

QY 66 THMPNRCYNATLCMHANTHYTGK-----INP-----SCPGIGCATV 104
Db 109 SQMHSSCYFSQCTCGNNTYFTAILQRTKSTSETNPVTSGLQPHGVLAGCDGTGVKSV 168

QY 105 CWTYFTHSTMSDGGIGQAREQVKEAISQLRGHTSPYKGLVLSKLHETLRTHTL 164
Db 169 CNWQQAIPHVDGGGQDVAVELYVOKIELVIOQFPKLSYHPLARKPRGP-DIDAQM 227

QY 165 VSLFNTLTRLH--VSAQNPT---NCWML-----PLHF-----RPY 197
Db 228 LDILSAT---HQALNISPSLAQNCWCLNQCSTNPFLAPVNVISFNASQNNCTPSLPF 283

QY 198 ISIPVPEQ--WNNF--STEINTSVLVGPVLSNLEITHTSNLTCTVKFSNTIDTTSQCIRW 254
Db 284 RVQMPESQYVPCFKAQNNSEDPVGV--VANP-----VNCSSSN-----322

QY 255 VTPPTTRIVCLPSGIFVCGTS--AYHCLNGSSESMSCFSLPVPMTIYTEODLYNHVVPKP 313
Db 323 ---HSEALCPGCGQAFVCGNNLFTALPANWTGSCVLAALLPDIDIDISGDD----PVPIP 375

QY 314 -----HNKRVPILPFVIRAGVLRGLTGTSITSTQFYKLSQIEINGMEQVTDLSL 365
Db 376 TFDYIAGRKRAVTLPLLVGLGVSTAVATGATAGLVAVQSYTKLSHQLINDVQALSSTI 435

QY 366 VTLQDOLNSLAAVLQNRALDILTAKEGTCGLFLGEERCYVYVNSRIVTEKVEIRDI 425
Db 436 NDLQDQVDSLAEVLLQNRGLDILTAEOQGICLALQERCFCYANKSGIVRDKIKNLQEDL 495

QY 426 QRAEQLQTERWGLLSQWMPVLPFLGFLAALILLLFGPCIFNLVVKVSSRIEAV 483
Db 496 EKRRKALADNLFLUTGLNGLLPYLLPFLPFAILLFSPFAPILRRVTLIRDLQNL 553

RESULT 6
B44282
retrovirus-related env polyprotein pseudogene - human
N;Alternate names: coat polyprotein
N;Contains: coat protein gp70; coat protein p20E
C;Species: Homo sapiens (man)
C;Date: 17-Feb-1994 #sequence_revision 04-Jan-1996 #text_change 09-Jul-2004
C;Accession: B44282
R;Hirose, Y.; Takamatsu, M.; Harada, F.
Virology 192, 52-61, 1993

A;Title: Presence of env genes in members of the RTVL-H family of human endogenous retroviruses
A;Reference number: A44282; MUID:932971138; PMID:8517031
A;Accession: B44282
A;Status: conceptual translation of pseudogene
A;Molecule type: mRNA
A;Residues: 1-570 <HIR>
A;Cross-references: UNIPROT:Q9N2K0; GB:D10083
C;Genetics:
C;Gene: env
C;Keywords: coat protein; polyprotein; pseudogene
F;1-372/Domain: coat protein gp70 <CGP>
F;370/Region: opal stop codon
F;373-569/Domain: coat protein p20E <CPP>
F;436-448/Region: immunosuppressive peptide

Query Match 18.1%; Score 476.5; DB 4; Length 570;
Best Local Similarity 34.7%; Pred. No. 1.3e-28;
Matches 119; Conservative 58; Mismatches 115; Indels 51; Gaps 13;

QY 189 CLPLHFRPYIS-----IPVPEQWNNFSTINTSVLVGLVSN---LEI 229
DB 221 CISH--PWLPSCSDSPRPSRLLIPSPEN-NSERLLIDTCFLIHENQNSPSMQL 277
QY 230 THTS---NLFCVKFSNTIDTSSQIRWV--TP-----PTRI-----VCLPSGIFVCGTS 275
DB 278 PQSPQLPLTAAALAGSLV-----WQDTFPTSPSHLTLHLQFCLTQGLFFLCGSS 330
QY 276 AYHCLNGSSSMCFSLFLVPPMTI--YTEQDLYNHVVPKHNKRVPLTPVIRAGVLG-- 331
DB 331 TYMCLPANWTGTCTLVLTPIKIOFANRTELPVLTPTKQKRVPLPIPLIPLMVLGUSAST 390
QY 332 -RLGTGIGSTTSTQFYKLSQEIINGDMQVDSLTVDLQNLAAVVLQNRRLDLT 390
DB 391 IALGTGAGISTVTFCSLNDFSASITDISQTLVLQVQVDSLAAVLQNRGLDLT 450
QY 391 AKRGCTCLFGEERCYYVNSRIVTEKVEIRDIQCRABELONTER--WGLLSQWMPWV 448
DB 451 AERKGGPCIFLNEECFYNGSLVYDNIKKLDRAQKLANQASNYAEPWPA--LSNMSWV 509
QY 449 LPFLGLAALILLGLGPCFNLIVKPVSSRIEAVKLQWVLM 491
DB 510 LPILSPLPIPLPLFPEFCFLVSLQIQNIAQITNLSTROM 552

RESULT 7
VCVDAR
env polyprotein - avian reticuloendotheliosis virus
N;Alternate names: coat polyprotein
N;Contains: coat protein gp22; coat protein gp73
C;Species: avian reticuloendotheliosis virus
C;Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 09-Jul-2004
C;Accession: A03999
R;Wilhelmsen, K.C.; Eggleston, K.; Temin, H.M.
J. Virol. 52, 172-182, 1984
A;Title: Nucleotide acid sequences of the oncogene v-rel in reticuloendotheliosis virus
A;Reference number: A93003; MUID:85009850; PMID:6090694
A;Accession: A03999
A;Molecule type: DNA
A;Residues: 1-582 <WIL>
A;Cross-references: UNIPROT:P03399
A;Experimental source: strain A
A;Note: strain A is a helper virus of the strain T
C;Comment: Enzymatic cleavages of env polyprotein may yield mature proteins including co
C;Genetics:
C;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; polyprotein

Query Match 17.6%; Score 463.5; DB 1; Length 582;
Best Local Similarity 28.3%; Pred. No. 1.3e-27;
Matches 150; Conservative 75; Mismatches 186; Indels 119; Gaps 20;

QY 9 LFTVLPPFALTAPPCCCTT-----SSSPQE-----FLXRTLPGNIDA 49

DB 96 VFKPIIPSVQBPQGECELTIIATQMHSTCYEKAQECTLLGKTYFTAILQKTLKLSYEDG 155
QY 50 PSYRSLSKGNSTFTTAHTHMRNCYNSATLCMHANTHYWTGKMINPSPCPGGLGATVCWTYF 109
DB 156 PN-----KLLQASCTGIWETSMGLGRCP-----CVC----- 181
QY 110 THTSMSDGGIQQAREKQVKAISQLTRGHSTPS--PYKGLVLSKLHETLRLTHRLVSLF 168
DB 182 ----LDGGPTDRF-GRICAEGLBEIIR-HSVPSVQYHPLALPR-----PRGVLDL 226
QY 169 NIT---LTRLHEV-SQNP---TNCWMLPLHFRPVIISIPVPEOWNNFSTEINTSVLVG 221
DB 227 POTSILEATHQVLAATNPQLAENCWLCMTLGTQS-----PQPSRRWAMS-----LSWEIA 277
QY 222 PLVSNLEITHSN---LTC-----VKFSNTIDTSSQICIRWWTPTTRIVCLP 265
DB 278 VLASLSGATHRVNRCQLLREADNRIGIPGVYVHFNCTSIQESLTERVIYEIRDYVLH 337
QY 266 SGIFFVCGTSAYHCLNGSSSMCFSLFLVPPMTIYTEQDLYNHVVPK-----HNKR 317
DB 338 RVMLVCEQHAHTALPNKWIIGLCILASIVPDMSIIPGEE-----PIPLPSIETAGRKRA 393
QY 318 VPILPVIIRAGVLRL---CTGIGSITSTQFYKLSQEIINGDMQVDSLTVDLQNL 374
DB 394 VQPIPLVGLIGITAGTAGGTGLG---VSVHTYHKLNSQLIEDVQALSGETINDLQDIDS 450
QY 375 LAAVVLQNRRLDLTLAKRGCTCLFGEERCYYVNSRIVTEKVEIRDIQCRABELON 434
DB 451 LAEVVLQNRGLDLTLTAQGGICLAEKCCFYANKSGIVRDKIRKLOEDLLARKALYD 510
QY 435 TERWGLLSQWMPWVLPFLGLAALILLGLGPCIFNLIVKPVSSRIEAVK 484
DB 511 NPLWNLGNGFLPYPVLLPSLGLFGLILFLTLGPCIRKTLTRIHDKIQGS 560

RESULT 8
S12812
env polyprotein - feline endogenous virus RD114 (fragment)
C;Species: feline endogenous virus RD114
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 20-Sep-1999
C;Accession: S12812
R;Moehring, R.
submitted to the EMBL Data Library, February 1990
A;Reference number: S12812
A;Accession: S12812
A;Molecule type: genomic RNA
A;Residues: 1-321 <MOE>
A;Cross-references: EMBL:X51930; NID:G61676; PIDN:CAA36196.1; PID:G61677
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: polyprotein

Query Match 15.8%; Score 415.5; DB 2; Length 321;
Best Local Similarity 31.2%; Pred. No. 2.8e-24;
Matches 96; Conservative 71; Mismatches 122; Indels 19; Gaps 6;

QY 198 ISIPVPEQWNNFSTINTSVLVGP--LVSNLEITHSNLTGVKFSNT-----IDTTS 248
DB 5 LAITPSTLTVSLADSLANASCOIIPVLLVQPMQFNSNCSLSPFINDTEQIDLGAVFTN 64
QY 249 SQCIRWWTPTTRIVCLPSGIFVCGTS--AYHCLNGSSSMCFSLFLVPPMTIYTEQD--- 304
DB 65 CTSVANVSSP---LCAALNGSVFLCGNNMAYTLPQNWTRLQVQASLLFDINDPGDEPVP 121
QY 305 --LYNHVVPKHNKRVPIPLPFVIRAGVIRGLRGIGSITSTQFYKLSQEIINGDMQV 362
DB 122 IPADHVIHRP-KRAVQFIPLLAGLIGITAAFTTGATGLGVSVTQYKLSHQLISDVQVLS 180
QY 363 DSLVTLQDQNLAAVVLQNRRLDLTLAKRGCTCLFGEERCYYVNSRIVTEKVEIR 422
DB 181 GTIQDLQDQVDSLAEVVLQNRGLDLTLTAQGGICLAEKCCFYANKSGIVRDKIRKLTQ 240
QY 423 DRICQRAEELQNTERWGLLSQWMPWVLPFLGLAALILLGLGPCIFNLIVKPVSSRIE 482

```
Db 241 EELQRRSLATNPLWTGLQGLFLLPLGLLTLTLITGCVFSRLMAFINDRLNV 300
Qy 483 VKLQMLVQ 490
Db 301 VHVMVLAQ 308

RESULT 9
T01381
env protein - murine leukemia virus
C:Species: murine leukemia virus
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C:Accession: T01381
R:Tumas, K.M.; Pozsgay, J.M.; Avidan, N.; Keiazek, S.J.; Overmeyer, B.; Blank, K.J.; Pry
Virology 192, 587-595, 1993
A:Title: Loss of antigenic epitopes as the result of env gene recombination in retroviri
A:Reference number: Z14313; MUID:93134803; PMID:7678475
A:Accession: T01381
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-676 <TM>
A:Cross-references: UNIPROT:O12374; EMBL:S53043; NID:9263651
C:Genetics:
A:Gene: env
C:Superfamily: type C retrovirus env polyprotein

Query Match 15.5%; Score 409; DB 2; Length 676;
Best Local Similarity 23.7%; Pred. No. 2.3e-23;
Matches 142; Conservative 81; Mismatches 219; Indels 158; Gaps 22;

Qy 16 PFALTAPPCCCTTS-----SSPYQELXRLTLPGN---IDAPSYRSLSKGNSTF 62
Db 97 PYSSAPGPPCCSGSGNIAGCARDNEPLTSLTPRCNTAGNRLKLDQVTHKS-SEGFYVC 155
Qy 63 TATHMTPR---NC-----YNSATLCMHANTHYWTGKINPSCGGLGATVCWTYFT-HT 112
Db 156 PG-SHREAPKSCGGPDSFYCASWGCETTGRTVYV---KPSSP---WDYIIVDN 201
Qy 113 SMSDGGIGIQGQAREK-----QVKEAISQLTRGH----- 140
Db 202 NLTSNQAVQCKONKNCNPLAQTINAGKQVTSWTTGHWGLRLYVSGDPLTGFGLRLS 261
Qy 141 -----STPSYKGLVLSKLHETL-----RTH 161
Db 262 YQNLGPRIPGNPVLADQLSFPLPNLPKPAKPPASSSTPTLISPSPTQPPVGTG 321
Qy 162 TRLVSLNTLTLRLHEVSAQNPTNCWMLPLHFRPYISIPVPEQWNNFSTEINTSVLVG 221
Db 322 DRLNLVQGAQALNLTNPDKTQECWLCVSGPPYEGVAV-----LGTYSNHTSA--- 372
Qy 222 PLVSNLEITHTSNLTCVKFSNTIDTSSQCIQRWTPPTPRIVC-----LPSG 267
Db 373 --PANCSVASQHKLTLSV-----TGRGLCLGTVPKTHQALCNTLTLAGKGSYLVVAPTG 425
Qy 268 IFFVCGTSAXHCLNGS-----SESMCFSLFVPPMTIYTEQDLYNHVVPKPHNKRVP1--- 320
Db 426 TMAACNTGLTCLSATVLNRTDVCVLVELWPRTYHPSPVYVQFEKSHRHKREPVSILT 485
Qy 321 LPFVIRAGVLRLGTGSGITTT-----STQFYKLSQEQINGMEQVTDLSVLTQDOLNSLAA 377
Db 486 LALLLGLTGGIAAGVGTGTALVATQFOQLHAAVQDDDLKEVEKSITNLEKSLTSLSE 545
Qy 378 VVLQNRALDLTAKRGTCCLPLGEBRCYVYNQSRIVTEKVEIRDRIQCRABELQNTER 437
Db 546 VVLQNRGLDILLFKEGGLCAALKECCFFADHTGLVDRSNMAKURELSQRKLFESSQG 605
Qy 438 WGLLSQW---PW-----VLPPLGLAALILLLLFGPCIFNLLVXFVSRIEAVKLQVLQ 490
Db 606 W--PEGFNRSFWTTLLISTIMGLPIIILLILLFGLPCILNRLVQFVKDRISVQALVLTQ 663
```

RESULT 10

```
VCVWEM
env polyprotein - Moloney murine leukemia virus
N:Contains: knob protein gp70; R protein; spike protein p15E
C:Species: Moloney murine leukemia virus
C:Date: 01-Sep-1981 #sequence_revision 24-Sep-1981 #text_change 09-Jul-2004
C:Accession: A93265; A93265; A93848; A03983
R:Shinnick, T.M.; Lerner, R.A.; Sutcliffe, J.G.
Nature 293, 543-548, 1981
A:Title: Nucleotide sequence of Moloney murine leukaemia virus.
A:Reference number: A93265; MUID:82035843; PMID:6169994
A:Accession: A93265
A:Molecule type: genomic RNA
A:Residues: 1-665 <SHI>
A:Cross-references: UNIPROT:P03385; GB:J02255; GB:J02256; GB:J02257; GB:M76668; NID:93311
A:Experimental source: clone pMLV-1
R:Sutcliffe, J.G.; Shinnick, T.M.; Green, N.; Liu, F.T.; Niman, H.L.; Lerner, R.A.
Nature 287, 801-805, 1980
A:Title: Chemical synthesis of a polypeptide predicted from nucleotide sequence allows d
A:Reference number: A93235; MUID:81052384; PMID:6159543
A:Accession: A93235
A:Molecule type: genomic RNA
A:Residues: 496-665 <SUT1>
A:Experimental source: provirus
R:Sutcliffe, J.G.; Shinnick, T.M.; Verma, I.M.; Lerner, R.A.
Proc. Natl. Acad. Sci. U.S.A. 77, 3302-3306, 1980
A:Title: Nucleotide sequence of Moloney leukemia virus: 3' end reveals details of replic
A:Reference number: A93848; MUID:81013872; PMID:6251454
A:Accession: A93848
A:Molecule type: DNA
A:Residues: 484-662, 'CEP' <SUT2>
A:Experimental source: provirus, clone pMLV-201
C:Genetics:
A:Gene: env
C:Superfamily: type C retrovirus env polyprotein
C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein
F:34-459/Product: knob protein gp70 #status predicted <KNB>
F:470-649/Product: spike protein p15E #status predicted <SPK>
F:650-665/Product: R protein #status predicted <RPT>
F:45,199,326,398,434/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 15.5%; Score 407.5; DB 1; Length 665;
Best Local Similarity 25.0%; Pred. No. 3e-23;
Matches 146; Conservative 77; Mismatches 227; Indels 133; Gaps 21;

Qy 16 PFALTAPPCCCTTS-----SPYQELXRLTLPGN---IDAPSYRSLSKGNSTF-- 62
Db 96 PFSSPPGPPCCSGSGSPGSRDCBEPILSLTPRCNTAARNRLKLDQVTHKS-----NEGFYV 151
Qy 63 TATHMTPR---NC-----YNSATLCMHANTHYWTGK-----MINPSCFGLGATVC-- 105
Db 152 CPGPHRPREKSCGGPDSFYCAYWGCETTGRTVYVYKPSSSWDFITVNNNLTSDQAVQCKD 211
Qy 106 -----WTYFTHTSMSDGG---GI----- 120
Db 212 NKWCNPLVIRFTAGRRVTSWTTGHWGLRLYVSGDPLTGFGLRLYQNLGPRVPIGPN 271
Qy 121 -----QGQAREKQVKEAISQLTRGHSTSPSYKGLVLSKLH-ETLRTHTRLVSLFNTLT 173
Db 272 PVLADQQLPKPKPVKSP-----SVTKPPSGTPLSPQLPPAGTENLLNLVDGAYQ 323
Qy 174 RLHEVSAQNPTNCWMLPLHFRPYISIPVPEQWNNFSTEINTSVLVGLVSNLEIT--- 230
Db 324 ALNLTSPDKTQECWLCVAGPPYVGVAVLGTYSNHTSAPANCASVASHQHLTLSEVTGQG 383
Qy 231 -----HTSNLTCVKFSNTIDTSSQCIQRWTPPTPRIVCLPSCIFFVCGTSAHCLNGS 283
Db 384 LCIGAVPKTHQALC-----NTQTSSRGSYYLVA-----PTGTMACSTGLTPCISTT 431
Qy 284 -----SESMCFSLFVPPMTIYTEQDLYNHVVPKPHNKRVP1---LPFVIRAGVLRLGTG 336
Db 432 ILLNLTDDYCVLVELWPRTYHPSPVYVQFEKSHRHKREPVSILTALLLGLTGGIAAG 491
Qy 337 IGSITT---STQFYKLSQEQINGMEQVTDLSVLTQDOLNSLAAVLQNRALDILLTAKR 393
```

Db 492 IGTGTTALMATQFOQLQAQVDDLRREVEKSIINLEKSLTSLSEVVLQNRGDLFLKE 551
Qy 394 GGTCLFLGEEBCYVYVQSRIVTEKVEIRDIQRAEELQNTERW--GLLSQMPW--- 447
Db 552 GGLCAALKKECCFYADHTGLVRDSMAKRLRNQKLFESTQWFEGLFNR-SPWFTTL 610
Qy 448 VLPFLGFLAALILLFGPCIFNLLVKFVSRSRIEAVKQWVLQ 490
Db 611 ISTIMGPLVLVLLMFLFGPCILNRLVQFVKDRISVVQALVLTQ 653

RESULT 11
VCVWFS
env polyprotein - mink cell focus-forming virus
N;Alternate names: coat polyprotein
N;Contains: knob protein gp70; R protein; spike protein p15E
C;Species: mink cell focus-forming virus
C;Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 23-Aug-1997
C;Accession: A31668
R;Chattopadhyay, S.K.; Baroudy, B.M.; Holmes, K.L.; Fredrickson, T.N.; Lander, M.R.; Mor
Virology 168, 90-100, 1989
A;Title: Biologic and molecular genetic characteristics of a unique MCF virus that is hi
A;Reference number: A31668; MUID:89085614; PMID:2535909
A;Accession: A31668
A;Molecule type: DNA
A;Residues: 1-636 <CHA>
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-440/Product: knob protein gp70 #status predicted <GP7>
F;441-620/Product: spike protein p15E #status predicted <P1E>
F;621-636/Product: R protein #status predicted <RPT>
F;43,58,297,329,369,405/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 15.3%; Score 402.5; DB 1; Length 636;
Best Local Similarity 25.8%; Pred. No. 6.8e-23;
Matches 132; Conservative 81; Mismatches 188; Indels 111; Gaps 18;
Qy 54 SLKSGNSTFTAHMTPRN---CYNATLCMHANTHYWTKMNPSPCGG----- 99
Db 149 SLKRGNT-----PRNQPCYDSSVV-----SSGIQATPGRCNPLVLEFTD 190
Qy 100 LGATVCW-----TYFHTSMSDGG-----IQGO---AREK 127
Db 191 AGKASWDGPKVWGLRLYRGTGIDPVTRFSLTRQVLNIGPRLPIGNPNVITGQLPSPRPV 250
Qy 128 QVKEAISQLRGHSTSPYKGLVLSKL---HETLRTHRLVSLFNLTTLRLHEVSAQNPT 184
Db 251 QIR-----LPRPQPPPPGAASIVPETAPSPQPGTGDRLLNLVDGAYQALNLTSPDKTQ 305
Qy 185 NCWMLCLPHRPRYISIVPQMNPFSTEINTSVLGPLVSNLEIT-----HTSN 234
Db 306 ECWLCLVAGPPYEGVAVLGTYSNHTSAPANCVASQHKLTLSVETGQGLCVGAVPKTHQ 365
Qy 235 LTCVKESNTIDTSSQICRWVTPTRIVCLPSGIFFCVCGTSAYHCLNGS-----SESMCFL 290
Db 366 ALC-----NTTQKTS-----GSYYLAAPAGTIWACNTGLTPCLSTTVLNITDYCVL 413
Qy 291 SFLVPPMTIYTEQDLNVHVPKPNKRVP1---LPFVIRAGVLGRGTGIGSITT---ST 344
Db 414 VELWPKVYHSPDYVYVTFEPGARFREPVSLLTALLGGLTWGGIAAGVGTGTALVAT 473
Qy 345 QFYFKLSQIEINGMEQVTDLSVLTQDLNSLAADVONRRALDILLTANRGGTCLFLGEER 404
Db 474 QFQOQLAAVHNDLKEVEKSITNLEKSLTSLSEVALQNRRLDILLFLKEGGLCAALKEEC 533
Qy 405 CYVYNQSRIVTEKVEIRDIQRAEELQNTERW--GLLSQMPW---VLPFLGFLAAL 458
Db 534 CFYADHTGLVRDSMAKRLRNQKLFESQGWFEGLFNR-SPWFTTLISTIMGPLVL 592

Qy 459 ILLLLFGPCIFNLLVKFVSRSRIEAVKQWVLQ 490
Db 593 LLILLFGPCILNRLVQFVKDRISVVQALVLTQ 624

RESULT 12
A43491
env polyprotein - T3651/B murine leukemia virus (strain T1223/B)
C;Species: T3651/B murine leukemia virus
C;Date: 12-Jan-1993 #sequence_revision 12-Jan-1993 #text_change 09-Jul-2004
C;Accession: A43491
R;Baylac-Kalabokias, H.; Astier-Gin, T.; Moynet, D.; Hernould, M.; Mamoun, R.; Legrand, I
Virus Res. 18, 117-134, 1991
A;Title: A new leukemogenic retrovirus isolated from tumor cells derived from a radio-in
A;Reference number: A43491; MUID:91253260; PMID:2042396
A;Accession: A43491
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-688 <BAY>
A;Cross-references: UNIPROT:Q9Q9A5; GB:X59002
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein

Query Match 15.2%; Score 400; DB 2; Length 688;
Best Local Similarity 23.7%; Pred. No. 1.2e-22;
Matches 141; Conservative 84; Mismatches 215; Indels 156; Gaps 21;
Qy 16 PFALTAPPCCCTSSS-----PYQEFLLRTRLPQIDAPSYRSLSGNSTFTAH 65
Db 111 PFSPPGPPCCGSSDSTGCRDCEPLTSTPRCNTAWN-----RLKLSK-----VTH 160
Qy 66 THM-----PRCYNATLCMHANTHYM-----TGK 90
Db 161 AHNEGFYVCPGPHRPRWARSCGPSEFYCASCWCETTGRASWKPSSWDYITVSNLTS 220
Qy 91 MINPSPCG-----GLGATV-----CWTFYHTSMSDGGTGGQAREKQVKE 131
Db 221 QATPACKGNKWCNLSLTIRFTSFQKQATSWVTGHWGLRLYVSGHDPGLIFG-----IRL 274
Qy 132 AISOQ-----TRGHSTSPYKGLVLSKLHET-----LRTHRLVS 166
Db 275 KTDLGRVPIGNPNVLSDRRPPSRPRTRSPPSNSTETETPLTPEPPPPAGVENRLN 334
Qy 167 LFNNTTLRLHEVSAQNPTNCMCL-----PLH-----FRPYISIP-----VPEQNNFS 210
Db 335 LVKGAQALNLTSPDKTQECWLCVSGPPYEGVAVLGTYSNHTSAPANCVASQHKLT 394
Qy 211 TEINTSVLVGPLVSNLEITHSNTLCVXFSTIDTSSQICRWVTPTRIVCLPSGIF 270
Db 395 SEVTGQGLCIGAVPKTQV-----LCNTTQKTS-----GSYYLAAPTGTIW 436
Qy 271 VCGTSAYHCLNGS-----SESMCFLSFLVPPMTIYTEQDLNVHVPKPNKRVP1---LPF 323
Db 437 ACSTGLTPCIISTILNLTDDYCVLVELWPRVYHSPYVYHQPERRAKYKREPVSILTAL 496
Qy 324 VIRAGVLGRGTGIGSITT---STQFYKLSQIEINGMEQVTDLSVLTQDLNSLAADV 380
Db 497 LLGLTWGGIAAGVGTGTALVATQFQQLQAAMHDDLKEVEKSITNLEKSLTSLSEV 556
Qy 381 QNRRALDILLTAKRGGTCLFLGEEBCYVYVQSRIVTEKVEIRDIQRAEELQNTERW-- 438
Db 557 QNRRLGDLFLKEGGLCAALKEECFYADHTGLVRDSMAKRLRNQKLFESQGWFE 616
Qy 439 GLLSQMPW---VLPFLGFLAALILLFGPCIFNLLVKFVSRSRIEAVKQWVLQ 490
Db 617 GLFNK-SPWFTTLISTIMGPLVLILLILLFGPCILNRLVQFVKDRISVVQALVLTQ 671

RESULT 13
B43491
env polyprotein - T3651/B murine leukemia virus
C;Species: T3651/B murine leukemia virus
C;Date: 12-Jan-1993 #sequence_revision 12-Jan-1993 #text_change 09-Jul-2004

A: Cross-references: UNIPROT:P26803; GB:M93134; NID:G331898; PIDN:AAA46478.1; PID:G331901
R:Remington, M.P.; Hoffman, P.M.; Ruscetti, S.K.; Masuda, M.
Nucleic Acids Res. 20, 3249, 1992
A:Title: Complete nucleotide sequence of a neuropathogenic variant of Friend murine leukemia virus
A:Reference number: G35474; MUID:92319660; PMID:1620621
A:Accession: G35476
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-676 <REMB>
A:Cross-references: EMBL:M93134; NID:G331898; PIDN:AAA46478.1; PID:G331901
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992
C:Genetics:
A:Gene: env
C:Superfamily: type C retrovirus env polyprotein
C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein
F:1-35/Domain: signal sequence #status predicted <SIG>
F:36-676/Product: env polyprotein #status predicted <ENV>
F:36-619/Domain: extracellular #status predicted <EXT>
F:36-479/Product: knob protein gp76 #status predicted <KGP>
F:476-479/Region: cleavage processing #status predicted
F:480-659/Product: spike protein p15E #status predicted <SPP>
F:486-502/Region: hydrophobic #status predicted
F:620-636/Domain: transmembrane #status predicted <TM1>
F:637-676/Domain: intracellular #status predicted <INT>
F:660-676/Product: R protein #status predicted <RPT>
F:46,202,336,368,408,444/Binding site: carbohydrate (Asn) (covalent) #status predicted

	Query Match	14.9%	Score 393;	DB 1;	Length 676;		
	Best Local Similarity	24.1%	Pred. No. 3.9e-22;				
	Matches 142;	Conservative 85;	Mismatches 226;	Indels 136;	Gaps 22;		
Qy	16	PFALTAPPCCCTSSS-----	-----PYQEFLEXRTRLPGN---	IDAPSYRLSKGNSTF	62		
Db	97	PYSPPPPCCGSSGNRAGCARD	DEPLTSLTPRCNTAWNRLKLDQVTHK--	SSGGFYV	154		
Qy	63	TAHTHMPR---NC-----	YNSATLCMHANTHYWTGK-----	MINPSPCGGLGATVC--	105		
Db	155	CPGSHRPRKAKSCGGPDYF	CASMGCECTGGRAYKWPSSWDYITVDNNLT	ITNQAQVCKD	214		
Qy	106	-----	WTYFHTSMSDGG-----	GIOGAAREKQVKEAI---	133		
Db	215	NKWCNPLAIQFTNAGKQVTS	WIGHYWGRLVXSQDPGLTFGIRLKYQNLGPRVPIGN	274			
Qy	134	----SOLT--RGHSTSPYKGLV	LSKLHETL-----	RTHRLVLSLENTTLTR	174		
Db	275	PVLADQLSFPLPNLPKPAK	SPSASNSTPLISPSAPTQPPAGTGRDLLNLVQ	GAYQA	334		
Qy	175	LHEVSAQNPTNCWMLPLHFR	PYISIPVPEQNN---FSTEINTTSLVGLVPLVSNLEI	TH	231		
Db	335	LNLTPDKTQBCWLCCL-----	VSAPPYEGVAVLGTYSNHTS---	APANCSAGSOH	382		
Qy	232	TSNLTCTVKFSNTIDTSSQ	IRWTPPTPRIVC-----	LPSGIFVCCTSAY	277		
Db	383	KLITUSEV-----	TGQGLCIGTPKTHQALCNTLT	KTGKGSYVLVAPAGTMMACNTGLT	435		
Qy	278	HCLNGS-----	SESMCFLSFLVPPMTIY	TEQDLYNHWPKPHNKRVP	1---	LPFVIRAGVL	330
Db	436	PCLSATVNLRTDYCVLVEL	PRVTHYHPYSYVSQFEKSYRHKRPVSLT	LALLIGLTM	495		
Qy	331	GRLTGTGSGITF---STQFY	KLSQEIINGDMQVDSLVTLQDQNSLA	AAVLQNRALD	387		
Db	496	GGIAAGVGTGTALVATQ	FOQLHAAVQDDLKEVEKSI	TLSSEVVLQNRGLD	555		
Qy	388	LLTAKRGCTCLFLGE	RCYYNQSRIVTEKVEIRDR	TCRAEELQNTERW--	GLLSQNM	445	
Db	556	LLFLKEGGLCAALKE	ECFCYADHTGLVSDMSAKLERLT	QROKLFESSQGNFEGFLNR-S	614		
Qy	446	PW-----	VLPPFLGPLAALILL	LLFGPCIFENLLV	KVFSYSSRIE	AVKQLQVLQ	490
Db	615	PWFPTLTSTINGPL	ILLILLILFGPCINRLVQF	VKDRISV	QVALVITQ	663	

RESULT 15

VCWVCB

env polyprotein - Cas-Br-E murine leukemia virus
N;Contains: coat protein p15E; knob protein gp70 precursor
C;Species: Cas-Br-E murine leukemia virus
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 09-Jul-2004
C;Accession: B26103
R;Rassart, E.; Nelbach, L.; Jolicœur, P.
J. Virol. 60, 910-919, 1986
A;Title: Cas-Br-E murine leukemia virus: sequencing of the paralytogenic region of its g
kemic tissues.
A;Reference number: A26103; MUID: 87061215; PMID: 3023680
A;Accession: B26103
A;Molecule type: DNA
A;Residues: 1-661 <RAS>
A;Cross-references: UNIPROT:P08360; GB:M14702; NID:G332016; PIDN:AAA46512.1; PID:G332018
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein
F;1-33/Domain: signal sequence #status predicted <SIG>
F;34-465/Product: knob protein gp70 #status predicted <KOP>
F;466-661/Product: coat protein p15E #status predicted <PTE>
F;43,186,199,322,361,394,430,598/binding site: carbohydrate (Asn) (covalent) #status pre

Query Match 14.9%; Score 392.5; DB 1; Length 661;
Best Local Similarity 22.8%; Pred. No. 4.2e-22;
Matches 142; Conservative 76; Mismatches 193; Indels 211; Gaps 23;
Qy 15 PPALTAPPCC-----CT-----TSPPQEFLLXRTLPGNIDAPSVRLSKG 58
Db 93 PYSPPGPPCCSGDAGVGCARDCEPLTYSYP-----RCNTAW 133
Qy 59 NSTFTAH-TMPCRYNSATLCMHANTHYWTKMINPSCPGGLCATVC-----105
Db 134 NRLKLARVTHAPKEG---YICGSHPRWA-----RSC-GGLDAYCASWGCTTGAA 184
Qy 106 -----WTFYFTMSDGGIQQAREKQVKEA-----ISQLTRG 139
Db 185 WNPTSSWDYIT-----VSNLTSSQATKACKNNGWCNPLVIRTPGKGRATSWTTG 235
Qy 140 H-----140
Db 236 HFWGLRLYISGHDPLFTGLRLKVTDLGPRVFGPNVLSQDRPPRVPARPPPSASP 295
Qy 141 STPS--PYKGLVLSKHLTLRLVSLFNTTLRLHEVSAQNPTNCWMLPLHFRPYI 198
Db 296 STPIPPQQG-----TGRLLNLVQGAVALTLNMTDPTTQECWLCVSEPPYE 344
Qy 199 SIPVPEQNNFSTEINTSVLGLVSNLEITHFSNLTVCVKSNTIDTTSSQIRWV---255
Db 345 GVAVLREYTSHT-----APANCSSGSQHKLTSEV-----TGQGRCLGTVPKT 388
Qy 256 -----TPPT-----RIVCLPSGIFFCGTSVAHCLNGS-----SESMCFLSFLVPPMTIY 300
Db 389 HQALCNREPTVSGSNLYVAPEGLTWACSTGLTFCLESTTVNLNLTTCVVLVELWPKVTYH 448
Qy 301 TEQDLYNHVVPKPHNKRVPV---LPFVIRAGVLRLGTGIGSIIT---STQFYKLSQEI 354
Db 449 SPDVVTYQFEGARPREPVSLLTALLPEGLTWGGIAAGVGTGTALVATQQFQOLQAM 508
Qy 355 NGDMEQVTDLSVLTQDQLNSLAAVVLQNRALDLLTAKRGCTCLFLGEERCYYVNSRIV 414
Db 509 HNDLKEVKSITNLEKSLTSLSEVVLQNRRLDLLFLKEGGLCAALKECCFCFYADHTGLV 568
Qy 415 TEKVKEIRDRIOCAEQLQNTERW---GLISQMPW-----VLPFLGLPLAALLILLFGPCI 468
Db 569 RDSMAKLERLNQKLFESGQGFGLFNR-SPWFTTLISTINGPLIVILLILLFGPCI 627
Qy 469 FNLLVFKVSSRIEAVKLQWLQ 490
Db 628 LNRLVQFVKDRISVVQALVLTQ 649

RESULT 16

S70395
env polyprotein - Friend murine leukemia virus (strain FB29)
N;Contains: knob protein gp76; R protein; spike protein p15E
C;Species: Friend murine leukemia virus
A;Variety: strain FB29
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C;Accession: S70395
R;Perryman, S.; Nishio, J.; Chesebro, B.
Nucleic Acids Res. 19, 6950, 1991
A;Title: Complete nucleotide sequence of Friend murine leukemia virus, strain FB29.
A;Reference number: S70393; MUID: 92107687; PMID: 1762923
A;Accession: S70395
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: genomic RNA
A;Residues: 1-676 <PER>
A;Cross-references: UNIPROT:P26804; EMBL:Z11128; NID:G61547; PIDN:CAA77479.1; PID:G61549
A;Experimental source: strain FB29
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; polyprotein; spike protein; transmembrane protein

Query Match 14.9%; Score 392; DB 2; Length 676;
Best Local Similarity 23.3%; Pred. No. 4.7e-22;
Matches 140; Conservative 81; Mismatches 221; Indels 158; Gaps 22;
Qy 16 PPALTAPPCCCTTS-----SSPYQEFLLXRTLPGN---IDAPSVRLSKGNSTF 62
Db 97 PYSPPGPPCCSGSGNWAGCARDCEPLTSLTPRCNTAWNRLKLDQVTHKS-SEGFYVC 155
Qy 63 TATHMTPR---NC-----YNSATLCMHANTHYWTKMINPSCPGGLCATVCVTFYHT 112
Db 156 PG-SHREPREAKSGGPDFSYCASWGCTTGVRVW-----KPSS-----SMDYITVDN 201
Qy 113 SMDGGGIQG-----QAREKQVKEAISQLTRGH-----140
Db 202 NLTSNAOVQVCKDNKWCNPLAIRFTNAGKQVTSWTTGHYWGRLYVSGDPLTFGIRLS 261
Qy 141 -----STPSPYKGLVLSKHLHETL-----RTH 161
Db 262 YQNLGPRIPIGPNPVLADQLSFLPNPLPKPAKSPASSSTPTLISPSPTTQPPAGTG 321
Qy 162 TRLVSLFNTTLRLHEVSAQNPTNCWMLPLHFRPYIISIPVPEQNNFSTEINTSVLVG 221
Db 322 DRLLNLVQGAVALTLNMTDPTTQECWLCVSGPPYEGVAV-----LGTYSNHTSA---372
Qy 222 PLVSNLEITHFSNLTVCVKSNTIDTTSSQIRWVTPPTTRIVC-----LPSG 267
Db 373 --PANCSSVASQHKLTSEV-----TGRCLGTVPKTHQALCNTTLKAGKGYLVAVPTG 425
Qy 268 IFVFCGTSVAHCLNGS-----SESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPV---320
Db 426 TWACNTGLTPCLUSATVLANRTDYCVLVELWPRVTHPPSYVYSQEKSHRHKREPVS LT 485
Qy 321 LPFVIRAGVLRLGTGIGSIIT---STQFYKLSQEIINGDMEQVTDLSVLTQDQLNSLAA 377
Db 486 LALLGGLTWGGIAAGVGTGTALVATQQFQOLHAAVQDDLKEVKSITNLEKSLTSLSE 545
Qy 378 VVLQNRALDLLTAKRGCTCLFLGEERCYYVNSRIVTEKVKEIRDRIOCAEQLQNTER 437
Db 546 VVLQNRGLDLLFLKEGGLCAALKECCFCFYADHTGLVDRSMKLERLSQKLFESSQ 605
Qy 438 WGLLSQMW---PW-----VLPFLGLPLAALLILLFGPCIENLLVFKVSSRIEAVKLQWLQ 490
Db 606 W---FEGWFRNSPWFTTLISTINGPLIVILLILLFGPCILNRLVQFVKDRISVVQALVLTQ 663

RESULT 17

VCWVGF
env polyprotein - feline leukemia virus (strain Gardner-Arnstein)
N;Contains: knob protein gp70; R protein; spike protein p15E

A>Title: Localization of neutralizing regions of the envelope gene of feline leukemia virus
A:Reference number: A25982; MUID:87061257; PMID:2431166
A:Accession: A25982
A:Molecule type: protein
A:Residues: 1-662 <ELD>
A:Cross-references: UNIPROT:Q7LVY7
C:Superfamily: type C retrovirus env polyprotein
C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein

Query Match 14.9%; Score 391.5; DB 2; Length 662;
Best Local Similarity 27.2%; Pred. No. 5e-22;
Matches 116; Conservative 71; Mismatches 158; Indels 81; Gaps 15;

Qy 124 AREKQVEAKISQLTRGHS-----TP-----SPYKGLVLSKLHETLRTHLVSLFN 169
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 259 SRQSQIE---SRVTPHHSSQGNGTGGITLVNASIAPLSTPVPASPKRIGTGDRLLNLVQ 315
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 170 TTLRLFEVSAQNPTNCNMCLPLHPRPVISIVPPQWNNFSTEINTTSVLGVPLVSNLEI 229
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 316 GTYLALNATDNRKYDCWLCLVRPPPYEGTAI---LGNYSNQTNP----- 359
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 230 THSNTLCVKFSNTIDTTSSQ--CIRWVTPPTRIVC-----LPSGIFPVCG 273
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 360 --PSCLSIPOHKLIISEVSGQGLCIGTVPKTHOALCNCTQQGHTGAHYLAAPNGTYACN 417
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 274 TSAVHC-----LNGSSSMCFSLFVLPPWTIYTEQDLNVHVVPKHKKRVIPI-----L 321
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 418 TGLTPCISMALVNLTSD--FCVLLIELPRVTYHQPEVVYVTFHFAKAARFRREPISLTVALML 476
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 322 PFVIRAGVLGRGTGIGSIITSTOPYYKLSQINGMEQVTDLSVTLQDQLNSLAADVLO 381
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 477 GGLTVGGTAGVGTTKALIETAQP-RQLQAMWHDIIQALESISALEKSLSLSEVVLQ 535
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 382 NRRALDLLTKARGGTCTFLGEERCYYVYNQSRIVTEKVKEIRDRIQCABELQNTERWGLL 441
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 536 NRRGDILFLQEGGLCAALKKECCFYADHTGLVRDNMAKLERLKQRQLFDSSQGW--F 593
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 442 SQWM---PW-----VLPFGPLAALLILLFGPCINLLVKFVSRIEAV-----KLQM 487
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 594 EGWFNKSPWFTLLISSMGPLLILLILLFGPCILNRLVQFVKDRISVVAQLILTQQYQ 653
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 488 VLOMEP 493
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 654 IKQYDP 659
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 19
VCMVVV
env polyprotein precursor - Rauscher mink cell focus-forming virus
N:Contains: coat protein p12e; coat protein p15e; knob protein gp70
C:Species: Rauscher mink cell focus-forming virus
C>Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 09-Jul-2004
C:Accession: A03990
R:Vogt, M.; Haggblom, C.; Swift, S.; Haas, M.
J. Virol. 55, 184-192, 1985
A>Title: Envelope gene and long terminal repeat determine the different biological properties of two strains of murine leukemia virus
A:Reference number: A93011; MUID:85237696; PMID:4009793
A:Accession: A03990
A:Molecule type: DNA
A:Residues: 1-640 <VOG>
A:Cross-references: UNIPROT:P06445; GB:M10100; NID:g332068; PIDN:AAA46528.1; PID:g332070
C:Genetics:
A:Gene: env
C:Superfamily: type C retrovirus env polyprotein
C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein
F:1-32/Domain: signal sequence #status predicted <SIG>
F:33-443/Product: knob protein gp70 #status predicted <KBX>
F:444-640/Product: coat protein p15e #status predicted <pFE>
F:444-623/Product: coat protein p12e #status predicted <pTE>
F:43, 58, 300, 332, 339, 372, 408, 576/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 14.8%; Score 390.5; DB 1; Length 640;
Best Local Similarity 29.1%; Pred. No. 5.7e-22;

144 AHNGGFVFCGPHRPRWARSCGGPESFYCASGCGEITGRASWKPSSSWDYITVSNLTS 203

91 MINPSCPG-----GLGATV-----CWYTFHTSMSDGG---GIQQAREKQ 128

204 QATPVCKGNEWNCNLSLIRFTSFGQATSWVTGHWGLRLYVSGHDPGLIFGIRLKITDSG 263

129 VKRAIS-----QLTRGHSTSPYKGLVLKUHET-----LRTHRLVSLFNTTL 172

264 PRVPIGNPNVLSDRRPPSPRPRTSPSPPSNSTPTETPLTLPPEPPAGVENRLNLVKGAY 323

173 TRLHEVSAQNPTNCWMLPLHLHFRVISIPVPEQWNNFSTEINTTSLVGLPYSNLEIT-- 230

324 QALNLSDPDKTQECWCLVCGPPYIEGVAVLGTYSNHTSAPANCVASOHKLTLSVETGQ 383

231 -----HTSNLTVCVFSNTIDTTSQICIRVWTPPTTRIVCLPISGIFFCVGSYAVHCLNG 282

384 GLCIGAVPKTHOVL-----NTTQKTS-----GSYLAAPTGTWACSTGLTFCIST 431

283 S-----SSSMCLFSLVPFMTIYITQDLYNVHVVPKHNKRVPI---LPFVIRAGVLGRLGT 335

432 TILDLTADYCVLVELWPRVTYVHSPSYVHQFERRAKYKREPVSLSLALILGGLTMGGIAA 491

336 GIGSITT---STQFYKLSQIEINGDMQGVTDLSVLTQDLNSLAAVLQNRREALDLITAK 392

492 GVGTTGTALVATQFQQLQAAMHDDLKEVEKSIINLEKSLTSLSEVLQNRRLGLDLLFLK 551

393 RGGTCLPLGEERCYYVNSQIRVTVEKVEIRDIRIOCRAEELONTERW---GLLSQWMPW--- 447

552 EGGLEAALKEECFYADHTGLVRSMAKRLRELQORQLFSSQOQWFEGLFNK-SFWFTT 610

448 -VLFFLGPLAALILLLLFGPCIFNLVLKVFSSRIEAVKLOMWLQ 490

611 LISTIMGPLIILLILLFGPCILNRLVQFIKDRISVWQALVLITQ 654

RESULT 21

VCMWLB

env polyprotein - feline leukemia virus (strain lambda-B1) (fragment)

N;Contains: coat protein gp70; coat protein p15E

C;Species: feline leukemia virus

C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 09-Jul-2004

C;Accession: A27172

R;Nicolaissen-Strouss, K.; Kumar, H.P.M.; Fitting, T.; Grant, C.K.; Elder, J.H.

J. Virol. 61, 3410-3415, 1987

A;Title: Natural feline leukemia virus variant escapes neutralization by a monoclonal ant

A;Reference number: A27172; MUID:88036192; PMID:2444714

A;Accession: A27172

A;Molecule type: DNA

A;Residues: 1-662 <NIC>

A;Cross-references: UNIPROT:P11261; GB:J03449; NID:G323896; PIDN:AAA43048.1; PID:G323897

C;Genetics:

A;Gene: env

C;Superfamily: type C retrovirus env polyprotein

C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein

F;1-33/Domain: signal sequence #status predicted <SIG>

F;34-465/Product: coat protein gp70 #status predicted <GPS>

F;466-662/Product: coat protein p15E #status predicted <GPT>

F;43,58,286,322,327,351,354,430/Binding site: carbohydrate (Asn) (covalent) #status predi

Query Match 14.8%; Score 389.5; DB 1; Length 662;

Best Local Similarity 27.8%; Pred. No. 7.1e-22;

Matches 115; Conservative 73; Mismatches 169; Indels 57; Gaps 15;

QY 124 AREKQVEAISQLTRGHS-----TP-----SPYKGLVLKSLHETLRTHRLVSLFN 169

DB 259 SRSQIE---SRVPHHSQNGNGFGITLVNASIASTPTVTPASPGRIGTGNRLINLVQ 315

QY 170 TTLTRLHEVSAQNPTNCWMLPLHLHFRPVISIPVPEQWNNFSTEINTTSLVGLPYSNLEI 229

DB 316 GTYALNVNTPNKTQCVLCLVSRPPYEGTAV---LGNYSNOTNPPPSCLSDPQHKLTI 372

QY 230 THTSNLT-CVCKFSNTIDTTSQCIIRVWTV---PPTRIVCLPISGIFFCVGSYAVHC-----L 280

Db 373 SEVSGQSCSI---GTVPKTHQALCKTKQGHKGTHYLAAPSGTYWACNTGLTPCISMAVL 429

Qy 281 NGSESCMFLVPLPPMTIYTEQDLYNHVVPKPKNKVPI-----LPFVIRAGVLGRL 333

Db 430 NWTSD-FCVLIELPWRVYTHQPEYVYTHFDKTRLRREPISLTVAMLGGLTVGGIAAGV 488

Qy 334 GTGSGTITSTQFYKLSQEINGMEQVTDLSLTVLQDOLNSLAADVONRAEDLLTAKR 393

Db 489 GTGPKALLETAQF-GQIQMAWHTDIOALEESISALEKSLTSLSEVVLONRRGLDILFQ 547

Qy 394 GGTCLFLGEERCYYVNOISRIYTEKVEIRDRIQCRABELQONTERWGLLSQW---PW--- 447

Db 548 GGLCAALKERCCFVADHTGLVRDNMAKRLERLQKQQLFDSQQW---FEGFNKSPWFTT 605

Qy 448 -VLPFLGLAALILLFLGFCIFNLLVKFVSSRIEAV-----KLQWVLOMEP 493

Db 606 LISSINGPLAILLILLFGPCILNRLVQFVKDRISVVQALILTQYQIKQYDP 659

RESULT 22

VCVMW2

env polyprotein - Friend mink cell focus-forming virus

N;Contains: knob protein gp70; spike protein p15E

C;Species: Friend mink cell focus-forming virus

C;Date: 20-Sep-1984 #sequence_revision 20-Sep-1984 #text_change 16-Jul-1999

C;Accession: A03989

R;Koch, W.; Zimmermann, W.; Oliff, A.; Friedrich, R.

J. Virol. 49, 828-840, 1984

A;Title: Molecular analysis of the envelope gene and long terminal repeat of Friend mink

A;Reference number: A03989; MUID:84138778; PMID:6321768

A;Accession: A03989

A;Molecule type: DNA

A;Residues: 1-627 <KOC>

A;Cross-references: GB:M12528; NID:g331918; PIDN:AAA46483.1; PID:g331920

C;Genetics:

A;Gene: env

C;Superfamily: type C retrovirus env polyprotein

C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein

F;1-32/Domain: signal sequence #status predicted <SIG>

F;33-431/Product: knob protein gp70 #status predicted <KNB>

F;432-627/Product: spike protein p15E #status predicted <SPK>

F;43,58,288,320,360,364,396/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 14.7%; Score 386; DB 1; Length 627;

Best Local Similarity 25.2%; Pred. No. 1.2e-21;

Matches 129; Conservative 79; Mismatches 184; Indels 120; Gaps 17;

Qy 54 SLKGNSTFAHTMPRN---CYNATLCHMANTHYWTGWINPSCPQG----- 99

Db 149 SLKRGNT-----PRNQPCYDSSAV-----SSGIQATPGRCNPLVLEFTD 190

Qy 100 LGATVCWYTFHTSMSPGGGIQGO-----AREKOVKEAISOLTRGHS--- 141

Db 191 AGKASH-----DGPVKWGLRLYRSTGDPVTRFSITRQVLNIGRIPGPNPVI 241

Qy 142 -----TPSPYKGLVLSKLHETLHTRLVLSFNLTTLRLHESVAQNPT 184

Db 242 DQLPPSPVQIMLPRPQPSPTGAASIQP-----GTGDRLLNVGAYQALNLTSPDKTQ 296

Qy 185 NCWMLPLHFRPYSISIVPQWNNFSTEINTTSLVGLVLSNLEIT-----HTSN 234

Db 297 ECWLCVLSGPPYBEGVAVLGTYSNHTSAPANCSVASQHKLTLSVTVQGLCVGAVPKTHQ 356

Qy 235 LTCVKFNTIDTSSQCIRWTPTRIVCLPSGFFVCGTSAYHCLNGS-----SESMCF 290

Db 357 ALC-----NTQNTSD-----GSYVLAAPAGTIWACNTGLTPCLSTVNLNLTDCVL 404

Qy 291 SPLVPPMTIYTEQDLYNHVVPKPKNKVPI---LPFVIRAGVLGRLGTGIGSITT---ST 344

Db 405 VELWPKVYTHSPGVYQFERKTKYKEPVSLSLTALLGLLTMGGAAGVGTGTALVAT 464

Qy 345 QFYKLSQEINGMEQVTDLSLTVLQDOLNSLAADVONRAEDLLTAKRGTCLFLGEER 404

Db 465 QQFQQLHAAVODDLKEVEKSIITNLEKSLTSLSEVVLQNRRLGLDLLFLKERGLCAALKEEC 524

Qy 405 CYYVNOISRIYTEKVEIRDRIQCRABELQONTERW---GLLSQWMPW---VLPFLGLAAL 458

Db 525 CFVADHTGLVRDSMAKRLERLQKLFESSQCFGLFNR-SPWFTTLISTINGPLIIL 583

Qy 459 ILLILLFGPCIFNLLVKFVSSRIEAVKLQWVLO 490

Db 584 LLILLFGPCILNRLVQFVKDRISVVQALVLTQ 615

RESULT 23

VCMVFP

env polyprotein precursor (clone CFE-6) - feline leukemia virus (provirus)

N;Contains: coat protein gp70; coat protein p15E

C;Species: feline leukemia virus

A;Note: host *Felis silvestris catus* (domestic cat)

C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999

C;Accession: A31479

R;Kumar, D.V.; Berry, B.T.; Roy-Burman, P.

J. Virol. 63, 2379-2384, 1989

A;Title: Nucleotide sequence and distinctive characteristics of the env gene of endogeno

A;Reference number: A31479; MUID:89199802; PMID:2539525

A;Accession: A31479

A;Molecule type: DNA

A;Residues: 1-668 <KUM>

A;Cross-references: GB:M25425; NID:g163849; PIDN:AAA30809.1; PID:g163850

C;Genetics:

A;Gene: env

C;Superfamily: type C retrovirus env polyprotein

C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein

F;1-33/Domain: signal sequence #status predicted <SIG>

F;34-465/Product: coat protein gp70 #status predicted <GUP>

F;466-668/Product: coat protein p15E #status predicted <TPP>

F;43,58,286,322,327,351,354,394,410,430,597/Binding site: carbohydrate (Asn) (covalent)

Query Match 14.6%; Score 385; DB 1; Length 668;

Best Local Similarity 28.6%; Pred. No. 1.6e-21;

Matches 120; Conservative 72; Mismatches 147; Indels 80; Gaps 16;

Qy 111 HTSMSPGG--GIQQAAREKOVKEAISOLTRGHSTPPSPYKGLVLSKLHETLHTRLVSLF 168

Db 271 HHPQGNNGTPTGI-----TLVNASIAPLSTPVTPTASP-----KRIGTGNRLNLV 314

Qy 169 NTLTRLHESVAQNPTNCWMLPLHFRPYSISIVPQWNNFSTEINTTSLVGLVLSNLE 228

Db 315 QGYTLNLVNTNPNKTKDCMLCLVSRPPYYEGIAV---LGNYSNQTNP----- 359

Qy 229 IHTSNLTCVKFNTIDTSSQ---CIRWVTPPTPRIVC-----LPSGIFVVC 272

Db 360 ---PSCLSVPQHKLTISEVSGQGLCIATVPKTHQALCNKTKQGHGTHYLVAHPNGTYWAC 416

Qy 273 GTSAYHC-----LNGSESCMFLVLPMTIYTEQDLYNHVVPKPKNKVPIPLPVIRA 327

Db 417 NTGLTFCISMAVLNWTSD-FCVLTELPWPRITYHEPIYSHFENKPKRKPDIISLTV--A 473

Qy 328 GVLGRGLTGIGSITTS-----TQFYKLSQEINGMEQVTDLSLTVLQDOLNSLA 378

Db 474 LMLG--GITVGGMARNRDCGLLETAQFQLQWAMHTDIOALEESISALEKSLTSLSEV 531

Qy 379 VLQNRALDLLTAKRGTCLFLGEERCYYVNOISRIYTEKVEIRDRIQCRABELQONTERW 438

Db 532 VLQNRGLDILFLQEGGLCTALKEECFCVADHTGLVRDNMAKRLERLQK-QQLFDSQOD 590

Qy 439 GLLSQW---PW-----VLPFLGLAALILLFLGPCIENLLVKFVSSRIEAVKLQWVLO 490

Db 591 G-LEGFNKSPWFTTLISSINGPLIILLLFGPCILNRLVQFVKDRISVVQALVLTQ 648

RESULT 24

A46511

envelope protein - AKV murine leukemia virus

C;Species: AKV murine leukemia virus

C;Date: 18-Jun-1993 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C;Accession: A46511
R;Hayashi, H.; Matsubara, H.; Yokota, T.; Kuwabara, I.; Kanno, M.; Koseki, H.; Isono, K.
J. Immunol. 149, 1223-1229, 1992
A;Title: Molecular cloning and characterization of the gene encoding mouse melanoma anti-
A;Reference number: A46511; MUID:92364323; PMID:1390036
A;Accession: A46511
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-669 <HAY>
A;Cross-references: UNIPROT:Q909A5; UNIPROT:Q9024; UNIPROT:Q83382
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:110845, NCBIP:110846)
C;Superfamily: type C retrovirus env polyprotein

Query Match 14.6%; Score 385; DB 2; Length 669;
Best Local Similarity 24.3%; Pred. No. 1.6e-21;
Matches 142; Conservative 79; Mismatches 231; Indels 132; Gaps 19;

QY 16 PFALTAPPPCCCTSSS-----PYQEFLLXRLPGNIDAPSYRSLSKGNSTFTAH 65
DB 94 PFSPPPGPPCGSSDSTSGCRDCEPLTSYTPRCNTAWN-----RLKLSK-----VTH 143
QY 66 THM-----PRNCYNATLCMHANTHW-----TGK 90
DB 144 AHNEGfYVCPGPHRWRARSCGPFSCASWCETTGRASWKPSSWDYITVSNLTS 203
QY 91 MINPSCPG-----GLGATV-----CWTFYTHTMSDGG---GIGQAREKQ 128
DB 204 QATPVCKGNKWCNSLTRFTSFQKQATSWTGHWWGLRLVSGHDPGLIFGIRKIDTSG 263
QY 129 VKEAIS-----QLTRGHSTPSPYKGLVLSKHET-----LRTHRLVSLFNHTL 172
DB 264 PRVPIGNPVLSDRRPPSRPRTPSPPPSNSTPTETPLTLPPEPPAGVENRLNLVKGAY 323
QY 173 TRLHEVSAQNPNCMCLPLHFRYISIPVPEOWNFSTINTSVLVGLPVNLEIT-- 230
DB 324 QALNLTSPDKTQECWCLVSGPPYEGVAVLGTYSNHTSAPANGSVASQHKLTSLSEVTGQ 383
QY 231 -----HTSNLTVCVKFSNTIDTSSQCIRWVTPTRIVCLPSGIFVCGTSAYHCLNG 282
DB 384 GLCIGAVPKTHQVLC-----NTQKTSDSGSYHVA-----PGTWACSTGLTPCIST 431
QY 283 S-----SESMCFSLFVPPMTIYTEQDLYNHVVPKPHNKRVI---LPFVIRAGVLGRGT 335
DB 432 TILNLTDDYCVLVELWPRVTYHSPSYVYHQFERRAKVKREPVSUTLALLLGLTMMGIAA 491
QY 336 GIGSITT---STQFYKLSQBEINGMEQVTDLSVTLQDQLNSLAADVQLNRRALDILLTAK 392
DB 492 GVGTTTALVATQFOQLQAAMHDDLKEVEKSIITNLEKSLTSLSEVVLQNRRLDILLFLK 551
QY 393 RGTGCLFGERECYVNVQSRIVTEKVKIIRDIOCRAEELQNTERW--GLLSQWMPW--- 447
DB 552 EGGUCAAKECCCFYADHTGLVRDSMAKRLRSORQKLFESQGWFEGLFNK-SPWFTT 610
QY 448 -VLFPFLGLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQWVLQ 490
DB 611 LISTIMGPLIILLLLFGPCILLNRLVQFIKDRISVVQALVLTQ 654

RESULT 25
VCVMW1
env polyprotein - mink cell focus-forming virus
N;Contains: knob protein gp70; R protein; spike protein p15E
C;Species: mink cell focus-forming virus
C;Date: 20-Sep-1984 #sequence_revision 20-Sep-1984 #text_change 16-Jul-1999
R;Mark, G.B.; Rapp, U.R.
J. Virol. 49, 530-539, 1984
A;Title: Envelope gene sequence of two in vitro-generated mink cell focus-forming murine
A;Reference number: A03987; MUID:84115078; PMID:6319752
A;Accession: A03987
A;Molecule type: DNA

A;Residues: 1-640 <MAR>
A;Cross-references: GB:K02725; NID:g331614; PIDN:AAA46375.1; PID:g331616
A;Experimental source: clone CI-3
A;Accession: B03987
A;Molecule type: DNA
A;Residues: 1-314,543-640 <MA2>
A;Experimental source: clone CI-4
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-441/Product: knob protein gp70 #status predicted <KNB>
F;442-621/Product: spike protein #status predicted <SPK>
F;622-640/Product: R protein #status predicted <RPT>
F;43,58,297,329,336,369/Binding site: carbohydrate (Asn) #status predicted
Query Match 14.6%; Score 384; DB 1; Length 640;
Best Local Similarity 24.7%; Pred. No. 1.8e-21;
Matches 128; Conservative 80; Mismatches 187; Indels 124; Gaps 18;

QY 54 SLKSGNSTFTAHTHMPRN---CYNATLCMHANTHWYTKMINPSCPG----- 99
DB 149 SLKRGNT-----PRNQGPCYDSSAV-----SSDIKGATPGGRCNPLVLEFTD 190
QY 100 LGATVCTVYTFHTMSDGGIQQ-----QAREKQVKEAIS----- 134
DB 191 AGKKASW-----DGPKVWGLRLYRSTGDPVTRFSLTROVLNIGPRVPVIGPNVIT 241
QY 135 -----QLTRGHSTPSPYKGLVLSKL---HETLRTHRLVSLFNHTLRLHEVSA 180
DB 242 DQLPPSRPVQIMLP RPQP PPGAA SIVPETAPPSQQLGTGDRLLNLVNGAYQALNLTSP 301
QY 181 QNPNCMCLPLHFRYISIPVPEOWNFSTINTSVLV-----GPLVSNLE 228
DB 302 DKTQECWCLVAGPPYEGVAVLGTYSNHTSAPANGSVASQHKLTSLSGVAGRGLCIAAFP 361
QY 229 IHTSNLTVCVKFSNTIDTSSQCIRWVTPTRIVCLPSGIFVCGTSAYHCLNGS---S 284
DB 362 KTHQA-----LCNTQKTSDSGSYH-----LAAPAGTIWACNTGLTCLSTTVLDLT 407
QY 285 ESMCFSLFVPPMTIYTEQDLYNHVVPKPHNKRVI---LPFVIRAGVLGRGTGIGSI 340
DB 408 TDYCVLVELWPKYTHSPSYVYHQFERRAKVKREPVSUTLALLLGLTMMGIAAGVGTG 467
QY 341 TT---STQFYKLSQBEINGMEQVTDLSVTLQDQLNSLAADVQLNRRALDILLTAKRGTC 397
DB 468 TTALVATQFOQLQAAMHDDLKEVEKSIITNLEKSLTSLSEVVLQNRRLDILLFLKEGGLC 527
QY 398 LFLGERECYVNVQSRIVTEKVKIIRDIOCRAEELQNTERW--GLLSQWMPW---VLPF 451
DB 528 AALKECCCFYADHTGLVRDSMAKRLRSORQKLFESQGWFEGLFNK-SPWFTTLLISTI 586
QY 452 LGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQWVLQ 490
DB 587 MGPLIILLLLFGPCILLNRLVQFIKDRISVVQALVLTQ 625

RESULT 26
VCVMW2
env polyprotein precursor - feline sarcoma virus (strain GA)
N;Contains: coat protein gp70; coat protein p15E
C;Species: feline sarcoma virus
A;Note: host Felis silvestris catus (domestic cat)
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 12-Apr-1996
R;Guilhot, S.; Hampe, A.; D'Auriol, L.; Galibert, F.
Virology 161, 252-258, 1987
A;Title: Nucleotide sequence analysis of the LTRs and env genes of SM-PeSV and GA-PeSV.
A;Reference number: A33741; MUID:98044502; PMID:2823466
A;Accession: B33741
A;Molecule type: DNA
A;Residues: 1-642 <GUI>
A;Gene: env

Db	310	NKTKDCWCLVSRPPYEGIAI---LGNYSQNTWPP-----PSCLSLTPQH 351
Qy	241	SNITDITSSQ--CIRWVTPPTRIVC-----LPSGIFVCGTSAYHC----- 279
Db	352	KLTISEVSGGCLCIGTVPRTHQALCNKTQOQHTGAHYLAAPNGTYWACNTGLTPCISMAV 411
Qy	280	LNGSSESMCFLSFLVPPMWTIYEQDLNHNVPKPHNKRVPI-----LPFVIRAGVLGR 332
Db	412	LNMTSD--FCVLIELWPRVTYHQPEYIYTHFDKAVRFRPEISLTVALMLGGTLVCGIAG 470
Qy	333	LGTGIGSTTSTQFYVKLSQINGMEQVDTSLVTLDQNLNSAAVLQNRRLDILLTAK 392
Db	471	VGTTKALLETAQF-RQLQIAHMTDIOALESISALEKSLTSLSEVVLQNRGLDILFLQ 529
Qy	393	RGGTCLFLGEERCYYVNSQSRIVTEKVKIEIRDIOCRABEELQNTERWGLLSQWM---PW-- 447
Db	530	GGGLCAALKEBCCFPADHTGLVRDNMAKRLERLQKQQLFDSQQGW--PEGWFNKSMPWT 587
Qy	448	--VLFPGLPLAALILLLLFGPCIFNLLVKFVSSRIEAV-----KLOMWLQMEP 493
Db	588	TLISSIMGPLILLILLFGPCILNRLVQFYKDRISVVVQAILTQYQOIQOYDP 642
RESULT 28		
TI0533		
env polyprotein precursor - feline leukemia virus (strain FelV-FAIDS)		
N;Contains: env protein gp70; env protein p15S		
C;Species: feline leukemia virus		
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004		
C;Accession: TI0533		
F;Donahue, P.R.; Hoover, E.A.; Beltz, G.A.; Riedel, N.; Hirsch, V.M.; Overbaugh		
J. Virol. 62, 722-731, 1988		
A;Title: Strong sequence conservation among horizontally transmissible, minima		
A;Reference number: Z17078; MUID:88119207; PMID:2828667		
A;Accession: TI0533		
A;Status: preliminary; translated from GB/EMBL/DBJ		
A;Molecule type: mRNA		
A;Residues: 1-642 <DON>		
A;Cross-references: UNIPROT:Q85522; EMBL:M18247; NID:g323904; PIDN:AAA93093.1;		
C;Superfamily: type C retrovirus env polyprotein		
Query Match 14.5%; Score 382; DB 2; Length 642;		
Best Local Similarity 28.2%; Pred. NO. 2.6e-21;		
Matches 106; Conservative 62; Mismatches 144; Indels 64; Gaps 12		
Qy	160	THRLVSLFNTTLRLHEVSAQNPTNCWMLPLHFRPYISIPVPEQMNFSFEINTTSVL 219
Db	286	TGDRILNLVQGTYALNATDNPNTKDWCLVSRPPYEGTAI---LGNYSQNTWPP--- 339
Qy	220	VGPLVSNLEITHTGNLTCVKFSNTIDTSSQ--CIRWVTPPTRIVC----- 263
Db	340	-----PSCLSIPQHKLTISEVSGGCLCIGTVPKTHQALCNKTQOQHTGAHYLA 387
Qy	264	LPSGIFVCGTSAYHC-----LNGSSESMCFLSFLVPPMWTIYEQDLNHNVPKPHNKRV 318
Db	388	APNGTYWACNTGLTPCISMAVNLNMTSD--FCVLIELWPRVTYHQPEYIYTHFAKAVRFRPE 446
Qy	319	PI-----LPFVIRAGVLGRGLTGIGSITSTQFYVKLSQINGMEQVDTSLVTLDQ 371
Db	447	PISLTVALMUGGLTVGGIAAGVGTGTKALLETAQF-RQLQIAHMTDIOALESISALEKS 505
Qy	372	LNSLAAVVLQNRRLDILLTAKRGGTCLFLGBERCYVNSQSRIVTEKVKIEIRDIOCRABE 431
Db	506	LTSLSEVVLQNRGLDILFLQEGGLCAALKEBCCFPADHTGLVRDNMAKRLERLQKQQL 565
Qy	432	LQNTERWGLLSQWM---PW-----VLFPGLPLAALILLLLFGPCIFNLLVKFVSSRIEAV- 483
Db	566	FDSQQGW--PEGWFNRSMPWTTLISSIMGPLILLILLFGPCILNRLVQFYKDRISVVQ 623
Qy	484	-----KLOMWLQMEP 493
Db	624	ALITQYQYQIKQYDP 639

RESULT 29
VCMVSA
env polyprotein precursor - feline leukemia virus (strain Sarma)
N;Alternate names: coat polyprotein
N;Contains: knob protein gp70; spike protein p15E
C;Species: feline leukemia virus
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 09-Jul-2004
C;Accession: A29013
R;Riedel, N.; Hoover, E.A.; Gasper, P.W.; Nicolson, M.O.; Mullins, J.I.
J. Virol. 60, 242-250, 1986
A;Title: Molecular analysis and pathogenesis of the feline aplastic anemia retrovirus, feline leukemia virus (FELV)
A;Reference number: A29013; MUID:86308240; PMID:3018287
A;Accession: A29013
A;Molecule type: DNA
A;Residues: 1-639 <RLE>
A;Cross-references: UNIPROT:P06752; GB:M14331; NID:G323898; PIDN:AAA43049.1; PID:G323898
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein
F;1-33/Domain: signal sequence #status predicted <SIG>
F;34-42/Product: knob protein #status predicted <NOB>
F;443-639/Product: spike protein #status predicted <SPK>
F;35,43,58,299,304,328,331,387,407,575/Binding site: carbohydrate (Asn) (covalent) #status predicted <STAT>
Query Match 14.4%; Score 380.5; DB 1; Length 639;
Best Local Similarity 27.6%; Pred. No. 3.3e-21;
Matches 112; Conservative 69; Mismatches 160; Indels 65; Gaps 14;
QY 124 AREKQVKEAIS---QLTRGHSTSPYKGLVLSKLHETLRLVSLFNLTTLRLHEVS 179
DB 247 SRQSTKSKVTQRQIT--SSTPRSVASATMGF--KRIGTGDRILNLVQGTVALNATD 302
QY 180 AQNPTNCWMLPLHFRYIISIPPEQWNNFSTEINTSVLVGLVSLNLEITHISNLTGVK 239
DB 303 PNKTKDCWLCVSRPPYEGIAV---LGNYSNOTNPP-----PSCSLTPQ 344
QY 240 FSNITDITTSQ--CIRWVTPPTIVC-----LPSGIFPVCGTSAVHC--- 279
DB 345 HKLTISEVSGQLCIGVPTKHQALCKKTKQKHGHTYLAAPNGTYWACNGLTPCISMA 404
QY 280 -LNGSSSMCFSLFVPPMTIYTEQDLYNHVVPKPHNKRVP-----LPFVIRAGVLG 331
DB 405 VLNWTSDFCVLIELFRVYHQPEYIYTHFDKAVRFRPREPISLTVALMGLTVGGIAA 463
QY 332 RLGTGIGSITSTQFYKLSQEINGMEQVTDLSVTLQDQNSLAAVVLQNRALDLTLTA 391
DB 464 GVGTTGKALLETAQF-RQLQIAHTDIQALEESISALEKSLTSLSEVVLQNRRLGLDLFL 522
QY 392 KRGTCTCLFGEERCYVYVQSRIVTEKVEIRDRIOCRABEELQNTWGLLSQW---PW- 447
DB 523 QEGGLCAALKEECFCYADHTGLVRDNNWAKRERLQKQQLFDSQQGW--FEGWFKSPWF 580
QY 448 ---VLPFLGPIAALILLILFGPCIFNLLVVFSSRIEAVKLQWLQ 490
DB 581 TTLISSMGPLLILLILLGPCIILNRLVQPKDRISVQVALIITQ 626

RESULT 30
VCMVHL
env polyprotein precursor - HoMuLV murine leukemia virus
N;Contains: knob protein gp76; spike protein p15E
C;Species: HoMuLV murine leukemia virus
A;Note: host Mus hortulanus (European mouse)
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 09-Jul-2004
C;Accession: B32594
R;Voytek, P.; Kozak, C.A.
Virology 173, 58-67, 1989
A;Title: Nucleotide sequence and mode of transmission of the wild mouse ecotropic virus, HoMuLV
A;Reference number: A32594; MUID:90051094; PMID:2554579
A;Accession: B32594

A;Molecule type: DNA
A;Residues: 1-666 <VOY>
A;Cross-references: UNIPROT:P21436
A;Note: the authors translated the codon CTT for residue 451 as Pro
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-466/Product: knob protein gp76 #status predicted <KPG>
F;467-666/Product: spike protein p15E #status predicted <SPK>
F;473-489/Domain: transmembrane #status predicted <TM1>
F;607-623/Domain: transmembrane #status predicted <TM2>
F;42,197,290,324,356,363,431,599/Binding site: carbohydrate (Asn) (covalent) #status predicted <STAT>
Query Match 14.4%; Score 380.5; DB 1; Length 666;
Best Local Similarity 24.3%; Pred. No. 3.5e-21;
Matches 146; Conservative 83; Mismatches 206; Indels 165; Gaps 27;
QY 15 PPALTAPPCCCTTSSSPYQSFELXKTRLPG---NIDAP-SYRSLSKGNST-----F 62
DB 92 PYSKPPGPP-CCTDNNP-----PGSRDCNGELTY--LTPRCSTAMNRLKVL 138
QY 63 TAH-----THMP--RNC-----YNSATLCMHANTHYWTGK-----MINPSC 96
DB 139 TTHLNLQGVYCGPHRPHARNCGPDPFYCAHWGCETTGQAYWKPSSSWDYIRVSNNA 198
QY 97 PGGLGATVC-----WT-----YFHTMSMD 116
DB 199 SSSDATACKNNWCSPLAISFTDPCKRATSWTSGFTWGLRLYISGHPGLIFGVLKISD 258
QY 117 GGG-----IQGAREKQVKEA---ISQLTRGHSTPS-----PYKGLVLSKLHETL 158
DB 259 LGPRVPIGPNVLSEQRPPSQPEPARLPPSSNLTOG-GTSPAPTGPPOEG----- 307
QY 159 RHTRVLVSLFNTLTLEHVSQNPNTNCWMLPLHFRPYISIPVPEQWNNFSTEINTSV 218
DB 308 -TGDRLLDLVQGAQALNATSPDKTOECWLCVSSPPYEGVAVVGVPYSNHIT----- 359
QY 219 LVGLVSNLEITHTSNLTCTKFSNTIDTTSQCIRWVTPPTTRIVCLPS----- 266
DB 360 --APANCADSQHLTLSEVT-GKPLPRKGSQ-----DPPGPVQYHSGARQKYSLSGGS 411
QY 267 GIFFVCGTSAYHCLNGS-----SESMCFSLFVPPMTIYTEQDLYNHVVPKPHNKRVP-- 320
DB 412 GTWACNTGLTPTCLSTAVLNLTIDYCVLVELWFRVTYHSLDFVYRQVEGRTRYOREPVS 471
QY 321 -LPFVIRAGVLGRLGTGIGSITTS---TOFYKLSQEINGMEQVTDLSVTLQDQNSLA 376
DB 472 TLALLGGLTWGGIAAGVGTGTSALVKTQFQGLHAAIQADLKEVSSITNLEKSLTSL 531
QY 377 AVVLQNRALDLTLTAKRGGTCLFGEERCYVYVQSRIVTEKVEIRDRIOCRABEELQNT 436
DB 532 EVVLQNRRLGLDLFLFKGGLCAALKEECFCYADHTGLVRDMSMAKLERLNQRKLFQAG 591
QY 437 RW--GLLSQWMPW-----VLPFLGPIAALILLILFGPCIFNLLVVFSSRIEAVKLQWLQ 490
DB 592 GWFEGLFNR-SPWLTTLISTINGPLIILLILFGPCILNRLVQPKDRISVQVALIITQ 650

Search completed: January 22, 2005, 22:13:31
Job time : 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2005, 21:08:36 ; Search time 137 Seconds
(without alignments)
2070.509 Million cell updates/sec

Title: US-09-319-156b-10
Perfect score: 2634
Sequence: 1 MALPYHTFTVLLPPFALT.....KFVSSRIEAVKQLWQLMEP 493

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 02:.*
1: uniprot_sprot:.*
2: uniprot_trembl:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2613	99.2	542	2 Q991W9	multiple sc
2	2335	88.6	538	1 ENW1 GORGO	P61561 gorilla gor
3	2323	88.2	538	1 ENW1 PANTR	P61563 pan troglod
4	2314	87.9	538	1 ENW1 HUMAN	Q9uqf0 h herv-w 7q
5	2313	87.8	538	1 ENW1 HYLPI	P61562 hyllobates p
6	2310	87.7	538	1 ENW1 PONPY	P61584 pongo pygma
7	1528	58.0	319	2 Q96L59	Q96L59 homo sapien
8	1522	57.8	319	2 Q96L60	Q96L60 homo sapien
9	1515	57.5	319	2 Q96L61	Q96L61 homo sapien
10	932	35.4	180	2 Q96TB1	Q96TB1 homo sapien
11	870	33.0	180	2 Q96TB4	Q96TB4 homo sapien
12	851	32.3	180	2 Q96TB5	Q96TB5 homo sapien
13	842.5	32.0	179	2 Q96TB3	Q96TB3 homo sapien
14	830	31.5	180	2 Q96TB2	Q96TB2 homo sapien
15	818	31.1	180	2 Q96TB6	Q96TB6 homo sapien
16	812	30.8	180	2 Q96TB7	Q96TB7 homo sapien
17	797.5	30.3	179	2 Q96TB8	Q96TB8 homo sapien
18	588	22.3	586	2 Q9IGU2	Q9IGU2 fowlpox vir
19	580	22.0	563	1 ENV_BAEYM	P10269 baboon endo
20	567.5	21.5	580	2 Q9WPZ9	Q9WPZ9 simian retr
21	562.5	21.4	574	1 ENV_SRV2	P51515 simian retr
22	559.5	21.2	564	2 Q9854	Q9854 rd114 retro
23	559.5	21.2	567	1 ENV_AVISN	P31796 avian splee
24	556	21.1	587	1 ENV_SRV1	P04027 simian retr
25	554.5	21.1	580	2 Q9WQ03	Q9WQ03 simian retr
26	553.5	21.0	574	1 ENV_SRV2R	P51520 simian retr
27	546	20.7	586	1 ENV_MPMV	P07575 simian maso
28	546	20.7	586	2 AAC82575	AAC82575 simian ma
29	534.5	20.3	618	2 Q8B141	Q8B141 mus musculu
30	524.5	19.9	620	2 Q9GLF7	Q9GLF7 trichosurus
31	514.5	19.5	584	1 ENH1_HUMAN	Q9n2k0 h herv-h_2q

32	481.5	18.3	538	1 EFRI_CALJA	P61553 callithrix
33	477.5	18.1	575	1 ENV_SMRVH	P21412 squirrel mo
34	477	18.1	537	1 EFRI_MACFA	P61556 macaca fasc
35	463.5	17.6	582	1 ENV_AVIRE	P03399 avian retic
36	463	17.6	538	1 EFRI_HYIML	P61555 hyllobates m
37	460	17.5	538	1 EFRI_HUMAN	P60508 homo sapien
38	460	17.5	538	2 AAH68585	AAH68585 homo sapi
39	459.5	17.4	538	1 EFRI_PONPY	P61558 pongo pygma
40	459	17.4	538	1 EFRI_PANTR	P61557 pan troglod
41	454	17.2	538	1 EFRI_GORGO	P61554 gorilla gor
42	445.5	16.9	563	1 ENH2_HUMAN	Q9n3j9 homo sapien
43	434	16.5	555	1 ENH2_HUMAN	Q9n3j8 homo sapien
44	428.5	16.3	584	1 EFCL_HUMAN	P60507 homo sapien
45	421.5	16.0	584	1 EFCL_PANTR	Q8mb16 pan troglod

ALIGNMENTS

RESULT 1

Q991W9	Q991W9	PRELIMINARY;	PRT;	542 AA.
AC	Q991W9;			
DT	01-JUN-2001 (Tremblrel. 17, Created)			
DT	01-JUN-2001 (Tremblrel. 17, Last sequence update)			
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)			
DE	Recombinant envelope protein.			
GN	Name=env;			
OS	Multiple sclerosis associated retrovirus element.			
OC	Viruses; Retroid viruses; Retroviridae.			
OX	NCBI_TaxID=89382;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21425392; PubMed=11531410;			
RA	Perron H., Jovin-Marthe E., Michel M., Ounanian-Paraz A., Camelo S.,			
RA	Dumon A., Jolivet-Reynaud C., Marcel F., Souillet Y., Borel E.,			
RA	Gebuhrer L., Santoro L., Marcel S., Seigneurin J.M., Marche P.N.,			
RA	Lafor M.,			
RT	"Multiple sclerosis retrovirus particles and recombinant envelope			
RT	trigger an abnormal immune response in vitro, by inducing polyclonal			
RT	Vbeta16 T-lymphocyte activation.";			
RL	Virology 287:321-332(2001).			
DR	EMBL; AF331500; AAK18189.1; -.			
DR	HSSP; P03385; IMOF.			
DR	GO; GO:0019028; C:viral capsid; IEA.			
DR	GO; GO:0019031; C:viral envelope; IEA.			
DR	GO; GO:0005198; F:structural molecule activity; IEA.			
DR	InterPro; IPR002050; Env_polyprotein.			
DR	Pfam; PF00429; Tlv_coat; 1.			
KW	Envelope protein.			
SQ	SEQUENCE 542 AA; 60624 MW; AAF3A8322E03C9EE CRC64;			
Query Match 99.2%; Score 2613; DB 2; Length 542;				
Best Local Similarity 98.8%; Pred. No. 6.7e-206;				
Matches 487; Conservative 2; Mismatches 4; Indels 0; Gaps 0;				
QY	1 MALPYHTFTVLLPPFALTAPPCCCTSSSPYQBFELKXRLPGNIDAPSYRLSKGNS 60			
DB	1 MALPYHTFTVLLPPFALTAPPCCCTSSSPYQBFELKXRLPGNIDAPSYRLSKGNS 60			
QY	61 TFTAHTMPKNCVNSATLCMHANTHTWTGKINPSCPGGLGATVCWTYFTHTSMDSGGI 120			
DB	61 TFTAHTMPKNCVNSATLCMHANTHTWTGKINPSCPGGLGATVCWTYFTHTSMDSGGI 120			
QY	121 QGQARKQVKEASQLTRGHSTPSYKGLVLSKLETHTLRLVSLFNTTLRLHEVSA 180			
DB	121 QGQARKQVKEASQLTRGHSTPSYKGLVLSKLETHTLRLVSLFNTTLRLHEVSA 180			
QY	181 QNPTNCWMCILPHFRPYISIPVPEQWNNFSTEINTTSLVGLVSLNLEITHNTLCVKF 240			
DB	181 QNPTNCWMCILPHFRPYISIPVPEQWNNFSTEINTTSLVGLVSLNLEITHNTLCVKF 240			
QY	241 SNTIDTTSQCIRWVTPPTTRIVCLPSGIFPVCGTSAVHCLNGSSBSCFSLVPPMTIY 300			

Db 241 SNTIDTSSQIRWVTPPTTRIVCLPSGIFVCGTSAVHCLNGSSECMFLSFLVPPMTIY 300
 QY 301 TEQDLYNHVVPKPHKVPILPFVIRAGVLRGLGTGIGSITTSQFYKLSQEIINGDMEQ 360
 Db 301 TEQDLYNHVVPKPHKVPILPFVIRAGVLRGLGTGIGSITTSQFYKLSQEIINGDMEQ 360
 QY 361 VTDLSVLTQDLQSLAAVVLQNRALDLLTAKRGCTCLFLGEERCYYVNSRIVTEKVK 420
 Db 361 VTDLSVLTQDLQSLAAVVLQNRALDLLTAKRGCTCLFLGEERCYYVNSRIVTEKVK 420
 QY 421 IRDRIQRAEELQNTERWGLLSQMPWLPFLGPLAAIILLFLGPGCIFNLLVKFVSSRI 480
 Db 421 IRDRIQRAEELQNTERWGLLSQMPWLPFLGPLAAIILLFLGPGCIFNLLVKFVSSRI 480
 QY 481 EAVKLQVLOWEP 493
 Db 481 EAVKLQVLOWEP 493

RESULT 2
 ENW1 GORGO
 ID _ENW1 GORGO STANDARD; PRT; 538 AA.
 AC P61561;
 DT 05-JUL-2004 (Rel. 44, Created)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE ERV-W1 provirus ancestral Env polypeptide precursor (Envelope
 polypeptide) (ERVW1 envelope protein) (Syncytin) [Contains: Surface
 protein (SU); Transmembrane protein (TM)].
 GN Name=ERVW1;
 OS Gorilla gorilla gorilla (Lowland gorilla).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
 OX NCBI_TaxID=9595;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=14757826; DOI=10.1073/pnas.0305763101;
 RA Mallet F., Bouton O., Prudhomme S., Cheynet V., Oriol G., Bonnaud B.,
 RA Lucotte G., Duret L., Mandrand B.;
 RT "The endogenous retroviral locus ERVW1 is a bona fide gene involved
 in hominoid placental physiology.";
 RL Proc. Natl. Acad. Sci. U.S.A. 101:1731-1736 (2004).
 CC -|- FUNCTION: Retroviral envelope proteins mediate receptor
 recognition and membrane fusion during early infection. Endogenous
 envelope proteins may have kept, lost or modified their original
 function during evolution. This endogenous envelope protein has
 retained its original fusogenic properties. In vitro it fuses
 cells expressing the type D mammalian retrovirus receptors SLC1A4
 and SLC1A5.
 CC -|- FUNCTION: SU mediates receptor recognition (By similarity).
 CC -|- FUNCTION: TM anchors the envelope heterodimer to the viral
 membrane through one transmembrane domain. The other transmembrane
 domain, called fusion peptide, mediates fusion of the viral
 membrane with the target cell membrane (By similarity).
 CC -|- SUBUNIT: The surface (SU) and transmembrane (TM) proteins form a
 heterodimer. SU and TM are attached by noncovalent interactions or
 by a labile interchain disulfide bond (By similarity).
 CC -|- DOMAIN: The CSX-17 immunosuppressive domain is present in many
 retroviral envelope proteins. As a synthetic peptide, it inhibits
 immune function in vitro and in vivo (By similarity).
 CC -|- PTM: Specific enzymatic cleavages in vivo yield the mature SU and
 TM proteins (By similarity).
 CC -|- PM: The CSX motif is highly conserved across a broad range of
 retroviral envelope proteins. It is thought to participate to the
 formation of a labile disulfide bond possibly with the CX6CC motif
 present in the transmembrane protein (By similarity).
 CC -|- MISCELLANEOUS: Ortholog of the human HERV-W_7q21.1 envelope
 protein.
 CC -|- SIMILARITY: Belongs to the gamma type-C retroviral envelope
 protein family. ERV class-IW subfamily.
 CC -|- CAUTION: CSX-17 sequence does not match the minimal active
 consensus.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AV101588; AAM68167.1; -;
 CC EMBL; AV101589; AAM68168.1; -;
 DR HSP; P03385; IMOF.
 DR InterPro; IPR002050; Env_polyprotein.
 DR Pfam; PF00429; TLV coat; 1.
 KW Cleavage on pair of basic residues; Envelope protein; ERV;
 KW Glycoprotein; Polypeptide; Signal; Transmembrane;
 KW Transposable element.
 FT SIGNAL 1 20 Potential.
 FT CHAIN 21 538 HERV-W1 provirus ancestral Env
 polyprotein.
 FT CHAIN 21 317 Surface protein (By similarity).
 FT CHAIN 318 538 Transmembrane protein (By similarity).
 FT TRANSMEM 320 340 Fusion peptide (Potential).
 FT TRANSMEM 444 464 Potential.
 FT SITE 186 189 CXXC (By similarity).
 FT SITE 317 318 Cleavage (By similarity).
 FT SITE 380 396 CSX-17 (By similarity).
 FT SITE 397 405 CX6CC (By similarity).
 FT CARBOHYD 169 169 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 208 208 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 214 214 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 234 234 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 281 281 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 409 409 N-linked (GlcNAc...) (Potential).
 SQ SEQUENCE 538 AA; 59939 MW; 95D4393FEDE196F3 CRC64;
 Query Match 88.6%; Score 2335; DB 1; Length 538;
 Best Local Similarity 89.2%; Pred. No. 4.6e-183;
 Matches 439; Conservative 16; Mismatches 37; Indels 0; Gaps 0;
 QY 1 MALPYHTPLFTVLLPPFALTAPPCCCTSSSPYQBFRLXRLPGNIDAPSRSLKSGNS 60
 Db 1 MALPYHILLFTVLLPSPFTTAPPCCMTSSSPYQBFRLMRMRPGNIDAPSRFSKGP 60
 QY 61 TFTAHTMPCNYSATLQWANTHYWTKMINPSCPGILGATVCWTYTHTSMSDGGI 120
 Db 61 TFTAHTMPCNYSATLQWANTHYWTKMINPSCPGILGATVCWTYTHTSMSDGGV 120
 QY 121 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLTHRLVLSLFTTLTRLHEVSA 180
 Db 121 QGQAREKHVKEVISQLTRVHSTSSPYKGLDLSKLHETLTHRLVLSLFTTLTGLHEVSA 180
 QY 181 QNPNTCMCLPLHFRPYISIPYEQWNNFSTEINTSVLGLVLSNLEITHSNLTCVKF 240
 Db 181 QNPNTCMCLPLDFRPYISIPYEQWNNFSTEINTSVLGLVLSNLEITHSNLTCVKF 240
 QY 241 SNTIDTSSQIRWVTPPTTRIVCLPSGIFVCGTSAVHCLNGSSECMFLSFLVPPMTIY 300
 Db 241 SNTIDTSSQIRWVTPPTTRIVCLPSGIFVCGTSAVHCLNGSSECMFLSFLVPPMTIY 300
 QY 301 TEODLYNHVVPKPHKVPILPFVIRAGVLRGLGTGIGSITTSQFYKLSQEIINGDMEQ 360
 Db 301 TEODLYNHVVPKPHKVPILPFVIRAGVLRGLGTGIGSITTSQFYKLSQEIINGDMEQ 360
 QY 361 VTDLSVLTQDLQSLAAVVLQNRALDLLTAKRGCTCLFLGEERCYYVNSRIVTEKVK 420
 Db 361 VTDLSVLTQDLQSLAAVVLQNRALDLLTAKRGCTCLFLGEERCYYVNSRIVTEKVK 420
 QY 421 IRDRIQRAEELQNTERWGLLSQMPWLPFLGPLAAIILLFLGPGCIFNLLVKFVSSRI 480
 Db 421 IRDRIQRAEELQNTERWGLLSQMPWLPFLGPLAAIILLFLGPGCIFNLLVKFVSSRI 480
 QY 481 EAVKLQVLOWEP 492

```

Db      481 EAVKQMEPKQK 492
||||||| :|:
RESULT 3
ENWI_PANTR
ID ENWI_PANTR STANDARD; PRT; 538 AA.
AC P61563;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE ERV-W1 provirus ancestral Env polyprotein precursor (Envelope
DE polyprotein) (ERVW1 envelope protein) (Syncytin) [Contains: Surface
DE protein (SU); Transmembrane protein (TM)].
DE Name=ERVW1;
GN Pan troglodytes (Chimpanzee).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
SEQUENCE FROM N.A.
RP PubMed=14757826; DOI=10.1073/pnas.0305763101;
RA Mallet P., Bouton O., Prudhomme S., Cheynet V., Oriol G., Bonnaud B.,
RA Lucotte G., Duret L., Mandrand B.;
RT "The endogenous retroviral locus ERVW1 is a bona fide gene involved
RL in hominoid placental physiology.";
CC -I- FUNCTION: Retroviral envelope proteins mediate receptor
CC recognition and membrane fusion during early infection. Endogenous
CC envelope proteins may have kept, lost or modified their original
CC function during evolution. This endogenous envelope protein has
CC retained its original fusogenic properties. In vitro it fuses
CC cells expressing the type D mammalian retrovirus receptors SLC14A
CC and SLC145.
CC -I- FUNCTION: SU mediates receptor recognition (By similarity).
CC -I- FUNCTION: TM anchors the envelope heterodimer to the viral
CC membrane through one transmembrane domain. The other transmembrane
CC domain, called fusion peptide, mediates fusion of the viral
CC membrane with the target cell membrane (By similarity).
CC -I- SUBUNIT: The surface (SU) and transmembrane (TM) proteins form a
CC heterodimer. SU and TM are attached by noncovalent interactions or
CC by a labile interchain disulfide bond (By similarity).
CC -I- DOMAIN: The CKS-17 immunosuppressive domain is present in many
CC retroviral envelope proteins. As a synthetic peptide, it inhibits
CC immune function in vitro and in vivo (By similarity).
CC -I- PTM: Specific enzymatic cleavages in vivo yield the mature SU and
CC TM proteins (By similarity).
CC -I- PTM: The CXXC motif is highly conserved across a broad range of
CC retroviral envelope proteins. It is thought to participate to the
CC formation of a labile disulfide bond possibly with the CXXCC motif
CC present in the transmembrane protein (By similarity).
CC -I- MISCELLANEOUS: Ortholog of the human HERV-W_7q21.1 envelope
CC protein.
CC -I- SIMILARITY: Belongs to the gamma type-C retroviral envelope
CC protein family. ERV class-I W subfamily.
CC -I- CAUTION: CKS-17 sequence does not match the minimal active
CC consensus.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AY101586; AAM68165.1; -
CC DR EMBL; AY101587; AAM68166.1; -
CC DR HSSP; P03385; 1MOF.
CC DR InterPro; IPR002050; Env_polyprotein.
CC DR Pfam; PF00429; TLV_coat; 1.
CC KW Cleavage on pair of basic residues; Envelope protein; ERV;
CC Glycoprotein; Polyprotein; Signal; Transmembrane;

```

```

KW Transposable element.
FT SIGNAL 1 20 Potential.
FT CHAIN 21 538 HERV-W1 provirus ancestral Env
FT polyprotein.
FT CHAIN 21 317 Surface protein (By similarity).
FT CHAIN 318 538 Transmembrane protein (By similarity).
FT TRANSMEM 320 340 Fusion peptide (Potential).
FT TRANSMEM 444 464 Potential.
FT SITE 186 189 CXXC (By similarity).
FT SITE 317 318 Cleavage (By similarity).
FT SITE 380 396 CKS-17 (By similarity).
FT SITE 397 405 CXXCC (By similarity).
FT CARBOHYD 169 169 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 208 208 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 214 214 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 234 234 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 242 242 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 281 281 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 409 409 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 538 AA; 59981 MW; A5DB5C51EF33AB0C CRC64;

Query Match 88.2%; Score 2323; DB 1; Length 538;
Best Local Similarity 88.6%; Pred. No. 4.5e-182;
Matches 436; Conservative 18; Mismatches 38; Indels 0; Gaps 0;

QY 1 MALPYHTFTLVLLPPFALTAPPCCCTTSSSPYQBFLLXRTLPGNIDAPSYSLKSGNS 60
DB 1 MALPYHTFTLVLLPPFALTAPPCCCTTSSSPYQBFLLXRTLPGNIDAPSYSLKSGNS 60
QY 61 TFTAHTHPRNCYSATLCHWANTHYWTGMINPSCPGGLGATVCWTYFTHTMSDGGI 120
DB 61 TFTAHTHPRNCYSATLCHWANTHYWTGMINPSCPGGLGATVCWTYFTHTMSDGGI 120
QY 121 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKXKHETLRTHTRLVSLFNTTLRLHEVSA 180
DB 121 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKXKHETLRTHTRLVSLFNTTLRLHEVSA 180
QY 181 QNPTNCMCLPLHFRYISIPVPEQWNNFSTEINTSVLGVPLVSNLEITHNTSLTCVKF 240
DB 181 QNPTNCMCLPLHFRYISIPVPEQWNNFSTEINTSVLGVPLVSNLEITHNTSLTCVKF 240
QY 241 SNTDITTSQCIRHWTPPTTRIVCLPSGIPFVCTSAHYCLNGSSESCFLSFLVPPMTIY 300
DB 241 SNTDITTSQCIRHWTPPTTRIVCLPSGIPFVCTSAHYCLNGSSESCFLSFLVPPMTIY 300
QY 301 TEQDLYNHVVPKPNKRVPILPFVIRAGVLGRIGTGIGTSITTSQFYKLSQELNGDMEQ 360
DB 301 TEQDLYNHVVPKPNKRVPILPFVIRAGVLGRIGTGIGTSITTSQFYKLSQELNGDMEQ 360
QY 361 VTDLSVTLQDQLNSLAAVVLQNRRLDLLTAKRGCTCLFLGBCRCYVNGSRIVTEKVK 420
DB 361 VADSLSVTLQDQLNSLAAVVLQNRRLDLLTAKRGCTCLFLGBCRCYVNGSRIVTEKVK 420
QY 421 IRDRIQCRABELONTERWGLLSQMPWLPFLGPLAAILLLLPGCCIFNLLVKFVSSRI 480
DB 421 IRDRIQCRABELONTERWGLLSQMPWLPFLGPLAAILLLLPGCCIFNLLVKFVSSRI 480
QY 481 EAVKQMEPKQK 492
DB 481 EAVKQMEPKQK 492

RESULT 4
ENWI_HUMAN
ID ENWI_HUMAN STANDARD; PRT; 538 AA.
AC Q9UQF0; O95244; O95245; Q8NHV7; Q9NRZ2; Q9NZG3;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE HERV-W_7q21.1 provirus ancestral Env polyprotein precursor (Envelope
DE polyprotein) (ERVW-7q Envelope protein) (HERV-W envelope protein)
DE (Syncytin) (Syncytin 1) (Enverin) (Env-W) [Contains: Surface protein
DE (SU); Transmembrane protein (TM)].

```

GN Name=ERVWE1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A., AND VARIANT ASN-307.
RC TISSUE=Placenta;
RX MEDLINE=21985840; PubMed=11990458;
RA Alliel P.M., Perin J.-P., Goudou D., Bitoun M., Robert B., Rieger F.;
RT "The HERV-W/7q family in the human genome. Potential for protein
expression and gene regulation.";
RL Cell. Mol. Biol. 48:213-217(2002).
RN [2]
RN SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=99099005; PubMed=9882319;
RA Blond J.-L., Beseme F., Duret L., Bouton O., Bedin F., Perron H.,
RT Mandrand B., Mallet F.;
RT "Molecular characterization and placental expression of HERV-W, a new
human endogenous retrovirus family.";
RL J. Virol. 73:1175-1185(1999).
RN [3]
RN SEQUENCE FROM N.A., TISSUE SPECIFICITY, FUNCTION, AND VARIANT ASN-307.
RC TISSUE=Testis;
RX MEDLINE=20155476; PubMed=10693809; DOI=10.1038/35001608;
RA Sha M., Lee X., Li X.-P., Veidman G.M., Finnerty H., Racie L.,
RA Lavallie E., Tang X.-Y., Edouard P., Howes S., Keith J.C. Jr.,
RA McCoy J.M.;
RT "Syncytin is a captive retroviral envelope protein involved in human
placental morphogenesis.";
RL Nature 403:785-789(2000).
RN [4]
RN SEQUENCE FROM N.A.
RX MEDLINE=20284713; PubMed=10826480;
RA Voisset C., Bouton O., Bedin F., Duret L., Mandrand B., Mallet F.,
RA Paranhos-Baccala G.;
RT "Chromosomal distribution and coding capacity of the human endogenous
retrovirus HERV-W family.";
RL AIDS Res. Hum. Retroviruses 16:731-740(2000).
RN [5]
RN SEQUENCE FROM N.A.
RX MEDLINE=22737999; PubMed=12853948; DOI=10.1038/nature01782;
RA Hillier L.W., Fulton R.S., Fulton L.A., Graves T.A., Pepin K.H.,
RA Wagner-McPherson C., Layman D., Maas J., Jaeger S., Walker R.,
RA Wyllie K., Sekhon M., Becker M.C., O'Laughlin M.D., Schaller M.E.,
RA Fowell G.A., Delehaunty K.D., Miner T.L., Nash W.E., Cordes M., Du H.,
RA Sun H., Edwards J., Bradshaw-Cordum H., Ali J., Andrews S., Isak A.,
RA Sunbrunt A., Nguyen C., Du F., Lamar B., Courtney L., Kalicki J.,
RA Ozersky P., Bielicki L., Scott K., Holmes A., Harkins R., Harris A.,
RA Strong C.M., Hou S., Tomlinson C., Dauphin-Kohlberg S.,
RA Kozlowicz-Reilly A., Leonard S., Rohlfing T., Rock S.M.,
RA Tin-Wollam A.-M., Abbott A., Minx P., Maupin R., Strommatt C.,
RA Latreille P., Miller N., Johnson D., Murray J., Woessner J.P.,
RA Wendt M.C., Yang S.-P., Schultz B.R., Wallis J.W., Spieth J.,
RA Bieri T.A., Nelson J.O., Berkowicz N., Wohlmann P.E., Cook L.L.,
RA Hickenbotham M.T., Eldred J., Williams D., Bedell J.A., Mardis E.R.,
RA Clifton S.W., Chissoe S.L., Marra M.A., Raymond C., Haugen E.,
RA Gillett W., Zhou Y., James R., Phelps K., Iadonoto S., Hubb K.,
RA Simms E., Levy R., Clendenning J., Kaul R., Kent W.J., Furey T.S.,
RA Baerbach R.A., Brent M.R., Keibler E., Flieck P., Bork P., Suyama M.,
RA Bailey J.A., Portnoy M.E., Torrents D., Chinwalla A.T., Gish W.R.,
RA Eddy S.R., McPherson J.D., Olson M.V., Eichler E.E., Green E.D.,
RA Waterston R.H., Wilson R.K.;
RT "The DNA sequence of human chromosome 7.";
RL Nature 424:157-164(2003).
RN [6]
RN FUNCTION, AND TISSUE SPECIFICITY.
RX PubMed=10708449;
RA Blond J.-L., Lavillette D., Cheynet V., Bouton O., Oriol G.,
RA Chapel-Fernandes S., Mandrand B., Mallet F., Cosset F.L.;
RT "An envelope glycoprotein of the human endogenous retrovirus HERV-W is
expressed in the human placenta and fuses cells expressing the type D

RT mammalian retrovirus receptor.";
RL J. Virol. 74:3321-3329(2000).
RN [7]
RN FUNCTION.
RX PubMed=11238877; DOI=10.1128/JVI.75.7.3488-3489.2001;
RA An D.S., Xie Y.-M., Chen I.S.Y.;
RT "Envelope gene of the human endogenous retrovirus HERV-W encodes a
functional retrovirus envelope.";
RL J. Virol. 75:3488-3489(2001).
RN [8]
RN FUNCTION.
RX PubMed=11531410; DOI=10.1006/viro.2001.1045;
RA Perron H., Jouvin-Marche E., Michel M., Ounanian-Paraz A., Camelo S.,
RA Dumon A., Jolivet-Reynaud C., Marcel F., Souillet Y., Borel E.,
RA Gebuhrer L., Santoro L., Marcel S., Seigneurin J.M., Marche P.N.,
RA Lafon M.;
RT "Multiple sclerosis retrovirus particles and recombinant envelope
trigger an abnormal immune response in vitro, by inducing polyclonal
Vbetale T-lymphocyte activation.";
RL Virology 287:321-332(2001).
RN [9]
RN FUNCTION.
RX PubMed=12050356;
RA Lavillette D., Marin M., Ruggieri A., Mallet F., Cosset F.-L.,
RA Kabat D.;
RT "The envelope glycoprotein of human endogenous retrovirus type W uses
a divergent family of amino acid transporters/cell surface
receptors.";
RL J. Virol. 76:6442-6452(2002).
RN [10]
RN FUNCTION.
RX MEDLINE=22946171; PubMed=14557543; DOI=10.1073/pnas.2132646100;
RA Blaise S., de Parseval N., Benit L., Heidmann T.;
RT "Genomewide screening for fusogenic human endogenous retrovirus
envelopes identifies syncytin 2, a gene conserved on primate
evolution.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:13013-13018(2003).
RN [11]
RN FUNCTION.
RX PubMed=12664292; DOI=10.1007/s00705-002-0960-x;
RA Ponferrada V.G., Mauck B.S., Wooley D.P.;
RT "The envelope glycoprotein of human endogenous retrovirus HERV-W
induces cellular resistance to spleen necrosis virus.";
RL Arch. Virol. 148:659-675(2003).
RN [12]
RN TISSUE SPECIFICITY.
RX PubMed=12970426;
RA de Parseval N., Lazar V., Casella J.-F., Benit L., Heidmann T.;
RT "Survey of human genes of retroviral origin: identification and
transcriptome of the genes with coding capacity for complete envelope
proteins.";
RL J. Virol. 77:10414-10422(2003).
RN [13]
RN DEVELOPMENTAL STAGE.
RX PubMed=12620933; DOI=10.1095/biolreprod.102.013078;
RA Smallwood A., Papageorghiou A., Nicolaides K., Alley M.K.R., Jim A.,
RA Nargund G., Ojha K., Campbell S., Banerjee S.;
RT "Temporal regulation of the expression of syncytin (HERV-W),
maternally imprinted PEG10, and SCGE in human placenta.";
RL Biol. Reprod. 69:286-293(2003).
RN [14]
RN VARIANTS ALA-129; GLN-138; ASN-307 AND PHE-477.
RX PubMed=14757826; DOI=10.1073/pnas.0305763101;
RA Mallet F., Bouton O., Prudhomme S., Cheynet V., Oriol G., Bonnaud B.,
RA Lucotte G., Duret L., Mandrand B.;
RT "The endogenous retroviral locus ERVWE1 is a bona fide gene involved
in hominoid placental physiology.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:1731-1736(2004).
CC -f- FUNCTION: Retroviral envelope proteins mediate receptor
recognition and membrane fusion during early infection. Endogenous
envelope proteins may have kept, lost or modified their original
function during evolution. This endogenous envelope protein has
retained its original fusogenic properties. It can make

pseudotypes with HIV-1 virions and confer infectivity. It participates to trophectoderm fusion during placenta morphogenesis. In vitro it fuses cells expressing the type D mammalian retrovirus receptors SLA14 and SLA15. It can induce cellular resistance to spleen necrosis virus in vitro.

-1- FUNCTION: SU mediates the envelope heterodimer to the viral membrane through one transmembrane domain. The other transmembrane domain, called fusion peptide, mediates fusion of the viral membrane with the target cell membrane (By similarity).

-1- SUBUNIT: The surface (SU) and transmembrane (TM) proteins form a heterodimer. SU and TM are attached by noncovalent interactions or by a labile interchain disulfide bond (By similarity).

-1- TISSUE SPECIFICITY: Expressed at higher level in placenta. Expressed at lower level in adrenal, bone marrow, brain, breast, colon, kidney, ovary, prostate, skin, spleen, testis, thymus, thyroid and trachea.

-1- DEVELOPMENTAL STAGE: In placenta, detected at higher level during early pregnancy and at lower level during late pregnancy.

-1- DOMAIN: The CKS-17 immunosuppressive domain is present in many retroviral envelope proteins. As a synthetic peptide, it inhibits immune function in vitro and in vivo (By similarity).

-1- PTM: Specific enzymatic cleavages in vivo yield the mature SU and TM proteins (By similarity).

-1- PTM: The CXXC motif is highly conserved across a broad range of retroviral envelope proteins. It is thought to participate to the formation of a labile disulfide bond possibly with the CX6CC motif present in the transmembrane protein. Isomerization of the intersubunit disulfide bond to an SU intrachain disulfide bond is thought to occur upon receptor recognition in order to allow membrane fusion (By similarity).

-1- POLYMORPHISM: All variants have fusogenic properties.

-1- DISEASE: Viral particles or intracellular RNA of HERV-W family members have been detected in tissue from patients with multiple sclerosis or schizophrania.

-1- MISCELLANEOUS: Orthologs in Pan troglodytes, Gorilla gorilla, Pongo pygmaeus and Hylobates moloch.

-1- MISCELLANEOUS: HERV-W family subgenomic RNAs have been observed.

-1- MISCELLANEOUS: This provirus is intergenic, the closest flanking genes being ODAG and PEX1.

-1- SIMILARITY: Belongs to the gamma type-C retroviral envelope protein family. HERV class-I W subfamily.

-1- CAUTION: CKS-17 sequence does not match the minimal active consensus.

-1- DATABASE: NAME=Human Endogenous Retrovirus Database (HERVD);

Query Match 87.9%; Score 2314; DB 1; Length 538;
Best Local Similarity 88.4%; Pred. No. 2.5e-181;
Matches 435; Conservative 19; Mismatches 38; Indels 0; Gaps 0;

Qy 1 MALPYHTFLFTVLLPFPALTAPPCCCTTSSSPYQEFLLXTRLPNGNIDAPSYSLKSGNS 60
Db 1 MALPYHTFLFTVLLPFPALTAPPCCCTTSSSPYQEFLLXTRLPNGNIDAPSYSLKSGTP 60

Qy 61 TFTAHTMPRNCYNATLCHMANTHYTGMKINPSCPGGLGATVCTWYFTHTMSDGGGI 120
Db 61 TFTAHTMPRNCYNATLCHMANTHYTGMKINPSCPGGLGATVCTWYFTHTMSDGGGV 120

Qy 121 QGQAREKQVKEAISQLRGHTSPSPYKGLVSLKHLHTLTHTRVLSLNTLTFLHEVSA 180
Db 121 QGQAREKQVKEAISQLRGHTSPSPYKGLVSLKHLHTLTHTRVLSLNTLTFLHEVSA 180

Qy 181 QNPNTNCWCLPLHFRPYISIPVQWNNFSTEINTSVLGPLVSNLEIHTSNLTCKVF 240
Db 181 QNPNTNCWCLPLHFRPYISIPVQWNNFSTEINTSVLGPLVSNLEIHTSNLTCKVF 240

Qy 241 SNTIDTSSQICRWVTPPTTRIVCLPSGIFFCVGTSAHYCLNGSSSESCFLSFLVPPMTIY 300
Db 241 SNTYTINSQICRWVTPPTTRIVCLPSGIFFCVGTSAHYCLNGSSSESCFLSFLVPPMTIY 300

Qy 301 TEQDLYNHVDPKPNKRVPIPLFVIRAGVLRGLTGIGSITTSQFYKLSQELNGDMEQ 360
Db 301 TEQDLYSVIVSKPRNKRVPILFVIRAGVLRGLTGIGSITTSQFYKLSQELNGDMEQ 360

Qy 361 VTDSLVTLODLNLSAAVLQNRALDILLTAKSGTCLFLGERCCYVNSQIRIVTEKVK 420
Db 361 VADSLVTLODLNLSAAVLQNRALDILLTAKSGTCLFLGERCCYVNSQIRIVTEKVK 420

Qy 421 IRDRIQCRABELQNTERWGLLSQWMPWLPFLGPLAAILLLFLFGPCIFNLLKVFSSRI 480
Db 421 IRDRIQCRABELQNTERWGLLSQWMPWLPFLGPLAAILLLFLFGPCIFNLLKVFSSRI 480

Qy 481 EAVKLQWLOME 492
Db 481 EAVKLQWMEPKNQ 492

RESULT 5
ENVI_HYLP1 ID ENVI_HYLP1 STANDARD; PRT; 538 AA.
AC P61562;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DE 01-OCT-2004 (Rel. 45, Last annotation update)
DE ERV-W1 provirus ancestral Env polypeptide precursor (Envelope polypeptide) (ERVW1 envelope protein) (Syncytin) [Contains: Surface protein (SU); Transmembrane protein (TM)].
DE protein (SU); Transmembrane protein (TM)].
GN Name-ERVW1;
OS Hylobates pileatus (Pileated gibbon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.
OX NCBI_TaxID=9589;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14757826; DOI=10.1073/pnas.0305763101;
RA Mallet F., Bouton O., Prudhomme S., Cheynet V., Oriol G., Bonnaud B., Lucotte G., Duret L., Mandrand B.;
RT "The endogenous retroviral locus ERVW1 is a bona fide gene involved in hominoid placental physiology.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:1731-1736(2004).
CC -1- FUNCTION: Retroviral envelope proteins mediate receptor recognition and membrane fusion during early infection. Endogenous envelope proteins may have kept, lost or modified their original function during evolution. This endogenous envelope protein has retained its original fusogenic properties. In vitro it fuses cells expressing the type D mammalian retrovirus receptors SLA14 and SLA15.
CC -1- FUNCTION: SU mediates receptor recognition (By similarity).
CC -1- FUNCTION: TM anchors the envelope heterodimer to the viral membrane through one transmembrane domain. The other transmembrane domain, called fusion peptide, mediates fusion of the viral membrane with the target cell membrane (By similarity).
CC -1- SUBUNIT: The surface (SU) and transmembrane (TM) proteins form a heterodimer. SU and TM are attached by noncovalent interactions or by a labile interchain disulfide bond (By similarity).
CC -1- DOMAIN: The CKS-17 immunosuppressive domain is present in many retroviral envelope proteins. As a synthetic peptide, it inhibits immune function in vitro and in vivo (By similarity).
CC -1- PTM: Specific enzymatic cleavages in vivo yield the mature SU and TM proteins (By similarity).
CC -1- PTM: The CXXC motif is highly conserved across a broad range of retroviral envelope proteins. It is thought to participate to the formation of a labile disulfide bond possibly with the CX6CC motif present in the transmembrane protein (By similarity).
CC -1- MISCELLANEOUS: Ortholog of the human HERV-W_7q21.1 envelope protein.
CC -1- SIMILARITY: Belongs to the gamma type-C retroviral envelope protein family. ERV class-I W subfamily.
CC -1- CAUTION: CKS-17 sequence does not match the minimal active consensus.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

```

-----
EMBL; AV101592; AAM68171.1; -.
DR EMBL; AV101593; AAM68172.1; -.
DR HSP; P03385; IMOF.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Tlv_coat; 1.
KW Cleavage on pair of basic residues; Envelope protein; ERV;
KW Glycoprotein; Polyprotein; Signal; Transmembrane;
KW Transposable element.
FT SIGNAL 1 20 Potential.
FT CHAIN 21 538 HERV-W1 provirus ancestral Env
FT CHAIN 21 317 polyprotein.
FT CHAIN 21 317 Surface protein (By similarity).
FT TRANSMEM 318 538 Transmembrane protein (By similarity).
FT TRANSMEM 320 340 Fusion peptide (Potential).
FT TRANSMEM 444 464 Potential.
FT SITE 186 189 CXXC (By similarity).
FT SITE 317 318 Cleavage (By similarity).
FT SITE 380 396 CKS-17 (By similarity).
FT SITE 397 405 CK6CC (By similarity).
FT CARBOHYD 169 169 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 208 208 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 214 214 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 234 234 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 242 242 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 281 281 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 409 409 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 538 AA; 60100 MW; E833A5AF7AF66BD7 CRC64;

Query Match 87.8%; Score 2313; DB 1; Length 538;
Best Local Similarity 88.4%; Pred. No. 3e-181;
Matches 435; Conservative 18; Mismatches 39; Indels 0; Gaps 0;

Qy 1 MALPYHTFLTVLLPPALAPPCCCTTSSSPQEFKXTRLPNGIDAPYSLSKGN 60
Db 1 MALPYHIFLTVLLPSFTLAPPCCRCMTSSSPQEFKXTRLPNGIDAPYSLSKGN 60
Qy 61 TFTAHTMPNCVNSATLQWANTHYWTGKINPSCPGGAGATVCTWYFTHSMDSGGI 120
Db 61 TFTAHTMPNCTVNSATLQWANTHYWTGKINPSCPGGAGATVCTWYFTHSMDSGGV 120
Qy 121 QGQAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRTLRLVSLFNTTLRLHEVSA 180
Db 121 QDQAREKHVKEVISQLTQVHSTSSPYKGLDLKSLHETLRTLRLVSLFNTTLRLHEVSA 180
Qy 181 QNPTNCWMLPLHPRPVISIPVPEQWNNFSTEINTTSLVGLPLVSNLEIHTSNLTCVKF 240
Db 181 QNPTNCWMLPLDPRPVVISIPVPEQWNNFSTEINTTSLVGLPLVSNLEIHTSNLTCVKF 240
Qy 241 SNTIDTSSOCIRWVTPTRIVCLPSGIFVCGTSAVHCLNGSSESCFLSELVPPMTIY 300
Db 241 SNTIDTSSOCIRWVTPTRIVCLPSGIFVCGTSAVHCLNGSSESCFLSELVPPMTIY 300
Qy 301 TEQDLYNHVVKPKHNKVPILPFVIRAGVLGRGTGIGSTTTSTQFYKLSQBIINGDMEQ 360
Db 301 TEQDLYNVVSKPNKVPILPFVIRAGVLGRGTGIGSTTTSTQFYKLSQBIINGDMEQ 360
Qy 361 VTSVLTVLQDLQSLAAVIVQNRALDILLTAKGGTCLFLGEBRCYVYVNSRIVTEKVE 420
Db 361 VADSVLTVLQDLQSLAAVIVQNRALDILLTAKGGTCLFLGEBRCYVYVNSRIVTEKVE 420
Qy 421 IRDRIQRAELQNTERRWGLLSQWMPVLPFLGPAALILLFPGCIFNLLVKEVSSRI 480
Db 421 IRDRIQRAELNIGWGLLSQWMPVLPFLGPAALILLFPGCIFNLLVKEVSSRI 480
Qy 481 EAVKLQWLOWE 492
Db 481 EAVKLQWEPKWE 492

```

RESULT 6

```

ENW1_PONPY STANDARD; PRT; 538 AA.
AC ENW1_PONPY
P61564;
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE ERV-W1 provirus ancestral Env polyprotein precursor (Envelope
DE polyprotein) (ERVW1 envelope protein) (Syncytin) [Contains: Surface
DE protein (SU); Transmembrane protein (TM)].
GN Name=ERVW1;
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OC NCBI_TaxID=9600;
RN [1]
RP PUBMED=14757826; DOI=10.1073/pnas.0305763101;
RA Mallet F., Bouton O., Prudhomme S., Cheynet V., Oriol G., Bonnaud B.,
RA Lucotte G., Duret L., Mandrand B.;
RT "The endogenous retroviral locus ERVW1 is a bona fide gene involved
RT in hominoid placental physiology.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:1731-1736(2004).
CC -I- FUNCTION: Retroviral envelope proteins mediate receptor
CC recognition and membrane fusion during early infection. Endogenous
CC envelope proteins may have kept, lost or modified their original
CC function during evolution. This endogenous envelope protein has
CC retained its original fusogenic properties. In vitro it fuses
CC cells expressing the type D mammalian retrovirus receptors SLC1A4
CC and SLC1A5.
CC -I- FUNCTION: SU mediates receptor recognition (By similarity).
CC -I- FUNCTION: TM anchors the envelope heterodimer to the viral
CC membrane through one transmembrane domain. The other transmembrane
CC domain, called fusion peptide, mediates fusion of the viral
CC membrane with the target cell membrane (By similarity).
CC -I- SUBUNIT: The surface (SU) and transmembrane (TM) proteins form a
CC heterodimer. SU and TM are attached by noncovalent interactions or
CC by a labile interchain disulfide bond (By similarity).
CC -I- DOMAIN: The CKS-17 immunosuppressive domain is present in many
CC retroviral envelope proteins. As a synthetic peptide, it inhibits
CC immune function in vitro and in vivo (By similarity).
CC -I- PTM: Specific enzymatic cleavages in vivo yield the mature SU and
CC TM proteins (By similarity).
CC -I- PTM: The CXXC motif is highly conserved across a broad range of
CC retroviral envelope proteins. It is thought to participate to the
CC formation of a labile disulfide bond possibly with the CXXCC motif
CC present in the transmembrane protein (By similarity).
CC -I- MISCELLANEOUS: Ortholog of the human HERV-W_7q21.1 envelope
CC protein.
CC -I- SIMILARITY: Belongs to the gamma type-C retroviral envelope
CC protein family. ERV class-I W subfamily.
CC -I- CAUTION: CKS-17 sequence does not match the minimal active
CC consensus.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
```

```

EMBL; AV101590; AAM68169.1; -.
EMBL; AV101591; AAM68170.1; -.
DR HSP; P03385; IMOF.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Tlv_coat; 1.
KW Cleavage on pair of basic residues; Envelope protein; ERV;
KW Glycoprotein; Polyprotein; Signal; Transmembrane;
KW Transposable element.
FT SIGNAL 1 20 Potential.
FT CHAIN 21 538 HERV-W1 provirus ancestral Env
FT CHAIN 21 317 polyprotein.
FT CHAIN 21 317 Surface protein (By similarity).

```

```
FT CHAIN 318 538 Transmembrane protein (By similarity).
FT TRANSMEM 320 340 Fusion peptide (Potential).
FT SITE 444 464 Potential.
FT SITE 186 189 CXXC (By similarity).
FT SITE 317 318 Cleavage (By similarity).
FT SITE 380 396 CKS-17 (By similarity).
FT SITE 397 405 CXCC (By similarity).
FT CARBOHYD 169 169 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 208 208 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 214 214 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 234 234 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 242 242 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 245 245 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 281 281 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 409 409 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 538 AA; 53908 MW; AAL7EE08E6D62A99 CRC64;

Query Match 87.7%; Score 2310; DB 1; Length 538;
Best Local Similarity 87.8%; Pred. No. 5.3e-181;
Matches 432; Conservative 19; Mismatches 41; Indels 0; Gaps 0;

QY 1 MALPYHTFLFTVLLPPFALTAPPCCCTTSSSPYQBFELXRLPLPGNIDAPSYRSLSGKNS 60
DB 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQBFELXRLPLPGNIDAPSYRSLSGKTP 60
QY 61 TFTAHTMPRNCYSATLCHWANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSGGGI 120
DB 61 TFTAHTMPRNCYSATLCHWANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSGGGV 120
QY 121 QGOAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTLRLVSLFNTTLRLHEVSA 180
DB 121 QGOAREKHVKEVISQLTRGHSTPSPYKGLDLSKLHETLRTHTLRLVSLFNTALTGLHEVSA 180
QY 181 QNPTNCWMLPLHFRPVIISIPVPEQWNNFSTEINTTSLVGLPLVSNLEIHTHTSNLTCVKF 240
DB 181 QNPTNCWMLPLAFRPVSIIPVPEQWNNFSTEINTTSLVGLPLVSNLEIHTHTSNLTCVKF 240
QY 241 SNTIDTSSQICRWVTPPTTRIVCLPSGIFPVCGTSAYHCLNGSSSFCFLSFLVPPMTIY 300
DB 241 SNTTNTTNSQICRWVTPPTQVCLPSGIFPVCGTSAYRCLNGSSSFCFLSFLVPPMTIY 300
QY 301 TEQDLYNHVVPKPNKRVPIIPVIRAGVLGRGTGSGITTSQFYKLSQELNGDMEQ 360
DB 301 TEQDLYNVVSKPRNKRVPILPFWMAAGVLGALTGIGGITTSTQFYKLSQBELNGDME 360
QY 361 VTDLSVLTQDLNSLAQVQLNRRALDILLTAKRGCTCLFLGEBRCYVYVNSQRIVTEKVKE 420
DB 361 VADSLVLTQDLNSIAAVVLQNRALDILLTAKRGCTCLFLGEBCCYVYVNSQIGIVTEKVKE 420
QY 421 IRDRIQRAEELQNTERWGLLSQMPWVLPFLGPLAAILLLLFGPCIFNLLVKFVSSRI 480
DB 421 IRDRIQRAEELRNIGPWGLFSQMPWVLPFLGPLAAILLLLFGPCIFNLLVNFVSSRI 480
QY 481 EAVKLQWVLOME 492
DB 481 EAVKLQWEPKQ 492

RESULT 7
Q96L59 PRELIMINARY; PRT; 319 AA.
AC Q96L59;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Envelope protein (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
[1]
RN Jern P., Karlsson D., Blomberg J.;
RP SEQUENCE FROM N.A.
QY 1 MALPYHTFLFTVLLPPFALTAPPCCCTTSSSPYQBFELXRLPLPGNIDAPSYRSLSGKNS 60
DB 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQBFELXRLPLPGNIDAPSYRSLSGKTP 60
QY 61 TFTAHTMPRNCYSATLCHWANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSGGGI 120
DB 61 TFTAHTMPRNCYSATLCHWANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSGGGV 120
QY 121 QGOAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTLRLVSLFNTTLRLHEVSA 180
DB 121 QGOAREKHVKEVISQLTRGHSTPSPYKGLDLSKLHETLRTHTLRLVSLFNTALTGLHEVSA 180
RA Jern P., Karlsson D., Blomberg J.;
```

```
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY050301; AAL11495.1; -.
DR GO; GO:0019031; C:viral envelope; IEA.
KW Envelope protein.
FT NON TER 319
SQ SEQUENCE 319 AA; 35784 MW; 7D288CCCD1B68999 CRC64;

Query Match 58.0%; Score 1528; DB 2; Length 319;
Best Local Similarity 87.8%; Pred. No. 4.9e-117;
Matches 280; Conservative 8; Mismatches 31; Indels 0; Gaps 0;

QY 1 MALPYHTFLFTVLLPPFALTAPPCCCTTSSSPYQBFELXRLPLPGNIDAPSYRSLSGKNS 60
DB 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQBFELXRLPLPGNIDAPSYRSLSGKTP 60
QY 61 TFTAHTMPRNCYSATLCHWANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSGGGI 120
DB 61 TFTAHTMPRNCYSATLCHWANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSGGGV 120
QY 121 QGOAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTLRLVSLFNTTLRLHEVSA 180
DB 121 QGOAREKHVKEVISQLTRGHSTPSPYKGLDLSKLHETLRTHTLRLVSLFNTALTGLHEVSA 180
QY 181 QNPTNCWMLPLHFRPVIISIPVPEQWNNFSTEINTTSLVGLPLVSNLEIHTHTSNLTCVKF 240
DB 181 QNPTNCWMLPLHFRPVIISIPVPEQWNNFSTEINTTSLVGLPLVSNLEIHTHTSNLTCVKF 240
QY 241 SNTIDTSSQICRWVTPPTTRIVCLPSGIFPVCGTSAYHCLNGSSSFCFLSFLVPPMTIY 300
DB 241 SNTIDTSSQICRWVTPPTTRIVCLPSGIFPVCGTSAYHCLNGSSSFCFLSFLVPPMTIY 300
QY 301 TEQDLYNHVVPKPNKRVPIIPVIRAGVLGRGTGSGITTSQFYKLSQELNGDMEQ 360
DB 301 TEQDLYNVVSKPRNKRVP 319

RESULT 8
Q96L60 PRELIMINARY; PRT; 319 AA.
AC Q96L60;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Envelope protein (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
[1]
RN Jern P., Karlsson D., Blomberg J.;
RP SEQUENCE FROM N.A.
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY050300; AAL11494.1; -.
DR GO; GO:0019031; C:viral envelope; IEA.
KW Envelope protein.
FT NON TER 319
SQ SEQUENCE 319 AA; 35763 MW; 67822669D1B6899D CRC64;

Query Match 57.8%; Score 1522; DB 2; Length 319;
Best Local Similarity 87.5%; Pred. No. 1.5e-116;
Matches 279; Conservative 8; Mismatches 32; Indels 0; Gaps 0;

QY 1 MALPYHTFLFTVLLPPFALTAPPCCCTTSSSPYQBFELXRLPLPGNIDAPSYRSLSGKNS 60
DB 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQBFELXRLPLPGNIDAPSYRSLSGKTP 60
QY 61 TFTAHTMPRNCYSATLCHWANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSGGGI 120
DB 61 TFTAHTMPRNCYSATLCHWANTHYWTGKMINPSCPGGLGATVCWTYFTHTSMSGGGV 120
QY 121 QGOAREKQVKEAISQLTRGHSTPSPYKGLVLSKLHETLRTHTLRLVSLFNTTLRLHEVSA 180
DB 121 QGOAREKHVKEVISQLTRGHSTPSPYKGLDLSKLHETLRTHTLRLVSLFNTALTGLHEVSA 180
```



```
DR Pfam: PF00429; TLV_coat; 1.
KW Envelope protein.
FT NON_TER 1
FT NON_TER 180
SQ SEQUENCE 180 AA; 20048 MW; F2B7B2FAF7CE40C3 CRC64;

Query Match 33.0%; Score 870; DB 2; Length 180;
Best Local Similarity 92.8%; Pred. No. 2.8e-63;
Matches 167; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY 207 NNFSTEINTTSLVGLPLVSNLEITHTSNLTVCVKSNTIDTSSQCIRWVTPPTTRIVCLPS 266
DB 1 NNFSTEINTTSLVGLPLVSNLEITHTSNLTVCVKSNTIDTSSQCIRWVTPPTTRIVCLPS 60
QY 267 GIFFVCGTSAYHCLNGSSSMCFSLFVLPMTIYTEODLYNHYVVPKPKRVPILPFVIR 326
DB 61 GIFFVCGTSAYHCLNGSSSMCFSLFVLPMTIYTEODLYNHYVVPKPKRVPILPFVIR 120
QY 327 AGVLGRIGTGIGSITTSQFYKLSQELNGDMEQVTDLSVTLQDQLNSLAAVVLQNRAL 386
DB 121 AGVLGRIGTGIGSITTSQFYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRAL 180

RESULT 12
Q96TB5 PRELIMINARY; PRT; 180 AA.
AC Q96TB5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Envelope protein (Fragment).
GN Name=env;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21268768; PubMed=11375061;
RA Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization,
RT identification, and phylogeny.";
RL AIDS Res. Hum. Retroviruses 17:643-648 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Kim H.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051008; BAB47560.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; TLV_coat; 1.
KW Envelope protein.
FT NON_TER 1
FT NON_TER 180
SQ SEQUENCE 180 AA; 13648 MW; 8D2F33813F16090B CRC64;

Query Match 32.3%; Score 851; DB 2; Length 180;
Best Local Similarity 91.1%; Pred. No. 1e-61;
Matches 164; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

QY 207 NNFSTEINTTSLVGLPLVSNLEITHTSNLTVCVKSNTIDTSSQCIRWVTPPTTRIVCLPS 266
DB 1 NNFSTEINTTSLVGLPLVSNLEITHTSNLTVCVKSNTIYTNQCIRWVTPPTTRIVCLPS 60
QY 267 GIFFVCGTSAYHCLNGSSSMCFSLFVLPMTIYTEODLYNHYVVPKPKRVPILPFVIR 326
DB 61 GIFFVCGTSAYHCLNGSSSMCFSLFVLPMTIYTEODLYNHYVVPKPKRVPILPFVIR 120
QY 327 AGVLGRIGTGIGSITTSQFYKLSQELNGDMEQVTDLSVTLQDQLNSLAAVVLQNRAL 386
DB 121 AGVLGRIGTGIGSITTSQFYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRAL 180
```

```
RESULT 13
Q96TB3 PRELIMINARY; PRT; 179 AA.
AC Q96TB3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Envelope protein (Fragment).
GN Name=env;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21268768; PubMed=11375061;
RA Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization,
RT identification, and phylogeny.";
RL AIDS Res. Hum. Retroviruses 17:643-648 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Kim H.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051008; BAB47560.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; TLV_coat; 1.
KW Envelope protein.
FT NON_TER 1
FT NON_TER 179
SQ SEQUENCE 179 AA; 19471 MW; 70DF3E5658B8E667 CRC64;

Query Match 32.0%; Score 842.5; DB 2; Length 179;
Best Local Similarity 90.6%; Pred. No. 5.1e-61;
Matches 163; Conservative 6; Mismatches 10; Indels 1; Gaps 1;

QY 207 NNFSTEINTTSLVGLPLVSNLEITHTSNLTVCVKSNTIDTSSQCIRWVTPPTTRIVCLPS 266
DB 1 NNFSTEINTTSLVGLPLVSNLEITHTSNLTVCVKSNTIDTSSQCIRWVTPPTTRIVCLPS 59
QY 267 GIFFVCGTSAYHCLNGSSSMCFSLFVLPMTIYTEODLYNHYVVPKPKRVPILPFVIR 326
DB 60 GIFFVCGTSAYHCLNGSSSMCFSLFVLPMTIYTEODLYNHYVVPKPKRVPILPFVIR 119
QY 327 AGVLGRIGTGIGSITTSQFYKLSQELNGDMEQVTDLSVTLQDQLNSLAAVVLQNRAL 386
DB 120 AGVLGRIGTGIGSITTSQFYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRAL 179

RESULT 14
Q96TB2 PRELIMINARY; PRT; 180 AA.
AC Q96TB2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Envelope protein (Fragment).
GN Name=env;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21268768; PubMed=11375061;
RA Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization,
RT identification, and phylogeny.";
```

```

RL AIDS Res. Hum. Retroviruses 17:643-648(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Kim H.;
RA Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051009; BAB47561.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; TLV_coat; 1.
KW Envelope protein.
FT NON_TER 1
FT NON_TER 180
SQ SEQUENCE 180 AA; 19709 MW; 4C8F196C9C788497 CRC64;

Query Match 31.1%; Score 830; DB 2; Length 180;
Best Local Similarity 88.3%; Pred. No. 5.4e-60;
Matches 159; Conservative 7; Mismatches 14; Indels 0; Gaps 0;

QY 207 NNFSTEINTSVLVGPLVSNLEITHTSNLTVCVKSNTIDTSSQCIKRWVTPPTTRIVCLPS 266
Db 1 NNFSTEINTSVLVGPLVSNLEITHTSNLTVCVKSNTIDTSSQCIKRWVTPPTTRIVCLPS 60
QY 267 GIFFVCGTSAHYCHLNGSSSMCFSLFVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIR 326
Db 61 GIFFVCGTSAHYCHLNGSSSMCFSLFVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIR 120
QY 327 AGVLGRLGTGIGITSTQFYKLSOEINGDMQVTDLSLTLQDQLNSLAADVVLQNRAL 386
Db 121 AGVLGRLGTGIGITSTQFYKLSOEINGDMQVTDLSLTLQDQLNSLAADVVLQNRAL 180

RESULT 16
Q96TB7 PRELIMINARY; PRT; 180 AA.
AC Q96TB7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Envelope protein (Fragment).
GN Name=env;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21268768; PubMed=11375061;
RA Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization,
RT identification, and phylogeny.";
RL AIDS Res. Hum. Retroviruses 17:643-648(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Kim H.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB050999; BAB47556.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; TLV_coat; 1.
KW Envelope protein.
FT NON_TER 1
FT NON_TER 180
SQ SEQUENCE 180 AA; 19984 MW; 3414BCDD74A060BA CRC64;

Query Match 30.8%; Score 812; DB 2; Length 180;
Best Local Similarity 87.8%; Pred. No. 1.6e-58;
Matches 158; Conservative 6; Mismatches 16; Indels 0; Gaps 0;

QY 207 NNFSTEINTSVLVGPLVSNLEITHTSNLTVCVKSNTIDTSSQCIKRWVTPPTTRIVCLPS 266
Db 1 NNFSTEINTSVLVGPLVSNLEITHTSNLTVCVKSNTIDTSSQCIKRWVTPPTTRIVCLPS 60
QY 267 GIFFVCGTSAHYCHLNGSSSMCFSLFVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIR 326
Db 61 GIFFVCGTSAHYCHLNGSSSMCFSLFVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIR 120
QY 327 AGVLGRLGTGIGITSTQFYKLSOEINGDMQVTDLSLTLQDQLNSLAADVVLQNRAL 386
Db 121 AGVLGRLGTGIGITSTQFYKLSOEINGDMQVTDLSLTLQDQLNSLAADVVLQNRAL 180

RESULT 17
Q96TB8 PRELIMINARY; PRT; 179 AA.
AC Q96TB8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Envelope protein (Fragment).
GN Name=env;

```

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=21268768; PubMed=11375061;
 RA Kim H.-S., Lee W.-H.;
 RA "Human endogenous retrovirus HERV-W family: chromosomal localization,
 RT identification, and phylogeny."
 RL AIDS Res. Hum. Retroviruses 17:643-648(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kim H.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB050996; BAB47555.1; -;
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002050; Env_polyprotein.
 DR Pfam; PF00429; TLV_coat; 1.
 KW Envelope protein.
 FT NON TER 1 179
 FT TER 179 179
 SQ SEQUENCE 179 AA; 19228 MW; 3A9FACB6376E74 CRC64;
 Query Match 30.3%; Score 797.5; DB 2; Length 179;
 Best Local Similarity 87.8%; Pred. No. 2.5e-57;
 Matches 156; Conservative 6; Mismatches 15; Indels 1; Gaps 1;
 QY 207 NNFSTEINTSVLGVPLVSLNLEIHTSNTLCVKFSNTIDTSSQIRWTPPTTRIVCLPS 266
 DB 1 NNFSTEINTSVLGVPLVSLNLEIHTSNTLCVKFSNTIANSQCIRRVTPPTTRIVCLTS 60
 QY 267 GIFVCVCTSAHYCLNGSSSEMCFLSVLPVPMIYTEQDLYNHVVPKPKRVPILPFPVIR 326
 DB 61 GIFVCVCTSAHYCLNGSSSEMCFLSVLPVPMIYTEQDLYNHVVPKPKRVPILPFPVIG 120
 QY 327 AGVLGRIGTGIGSTTTSTQFYKLSQIEINGMEQVDSLVTLQDQLNSLAAVVLQNRAL 386
 DB 121 AGVLGLGTGGGITASTQF-YKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNTAL 179
 RESULT 18
 ID Q9IGU2 PRELIMINARY; PRT; 586 AA.
 AC Q9IGU2;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Envelope glycoprotein.
 OS Fowlpox virus (FPV).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Avipoxvirus.
 OX NCBI_TaxID=10261;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Singh P., Kim T.-J., Tripathy D.N.;
 RT "Re-emerging fowlpox: evaluation of isolates from vaccinated flocks."
 RL Avian Pathol. 29:449-455(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Singh P., Schitzlein W.M., Tripathy D.N.;
 RT "Reticuloendotheliosis Virus Sequences Within the Genomes of Field
 RT Strains of Fowlpox Virus Display Variability."
 RL J. Virol. 0:0-0(2003).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Schitzlein W.M., Singh P., Srinivasan V., Tripathy D.N.;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF246698; AAF81698.2; -;
 DR HSP; P03385; IMOF.
 DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR008965; Cellul_bind.
 DR InterPro; IPR002050; Env_polyprotein.
 DR Pfam; PF00429; TLV_coat; 1.
 KW Envelope protein.
 SQ SEQUENCE 586 AA; 63723 MW; DBB221FA9FC562C3 CRC64;
 Query Match 22.3%; Score 588; DB 2; Length 586;
 Best Local Similarity 32.6%; Pred. No. 1.9e-39;
 Matches 161; Conservative 74; Mismatches 161; Indels 98; Gaps 20;
 QY 63 TAHTMPRNCYSATLQCHANTHYWT-----GKMINPSCGGLGATVCWT 107
 DB 118 TIATQMHSTCYERAQECTLLGKTYFTAILQKTLGSLYEDGNKLLQASCTGTGKPCWD 177
 QY 108 YPHTSMGSGGIGQQAEREQVKEAISQLRGHSTPS-PYKGLVLSKLHETLRTHRLVS 166
 DB 178 PVAPVYVSDGGPTDMIRESVRELEIR-HSYPSVQYHPLALPR-----PRGVD 228
 QY 167 LFNTT---LTRLHEV-SAQNPN---TNCWMC-----LPLH 193
 DB 229 LDPQTSIDLEATHQVLNATNPQLAENCWLCWLTGTPIPAAIPANGNVTLDCNCSLSLPR 288
 QY 194 FRPVISIPVEQNNFSTEI-NTTSLVGLVPLVSLNLEIHTSNTLCVKFSNTIDTSSQCI 252
 DB 289 VQPTGSDV---NCYAGEADNRTGPIG-----YVHFTN--CTSIOEVSNETSH--I 333
 QY 253 RWTPTPTTRIVCLPSGIPFVCGTS-AVHCLNGSSSEMCFLSVLPVPMIYTEQDLYNHVVP 311
 DB 334 RNLTR---LCPPPGHVFVCGNNAYTALPNKVGILCILASIVPDMISIIGEE---PIP 385
 QY 312 KP-----HNKRVLPILPFIIRAGVLGR---GTGIGSTTTSTQFYKLSQIEINGMEQ 360
 DB 386 LPSIEYTAGRHKRAVQFIPLLVGLGITAGTAGTGLG---VSVHTYKLSNQLIEDVQA 442
 QY 361 VTDSLVTLDQLNSLAAVVLQNRALDILLTAKGGTCLFLGEBRCYVYVNSRIVTEKVE 420
 DB 443 LSGTINDLQDIDSLAEVVLQNRGLDILLTAEGGICLALQEKCCFYANKSGIVRDKIRK 502
 QY 421 IRDRIQCRABELQNTERWGLLSQMPWLPFLGPLAALILLLLFGPCIFNLLVKFVSRI 480
 DB 503 LQEDLIARKKALYDNPWNGINGFLPYLLPLGLPLGLILLFLTLGPCIITLRIIHDKI 562
 QY 481 EAVK-LQWLMQEP 493
 DB 563 QAVKILALVPQYKP 576
 RESULT 19
 ID ENV_BAEVM STANDARD; PRT; 563 AA.
 AC P10269;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Env polyprotein precursor (Coat polyprotein) [Contains: Outer membrane
 DE protein gp70; Transmembrane protein p20E].
 GN Name=ENV;
 OS Baboon endogenous virus (strain M7).
 OC Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses.
 OX NCBI_TaxID=11764;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kato S., Matsuo K., Nishimura N., Takahashi N., Takano T.;
 RT "The entire nucleotide sequence of baboon endogenous virus DNA: a
 RT chimeric genome structure of murine type C and simian type D
 RT retroviruses."
 RL Jpn. J. Genet. 62:127-137(1987).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/annouce/> or send an email to license@isb-sib.ch).

EMBL; D10032; BAA00924.1; --
EMBL; X05470; CAA29028.1; --
PIR; J0262; VCMVM7.
HSP; P03385; IMOF.
InterPro; IPR002050; Env_polyprotein.
Pfam; PF00429; TLV_coat; 1.
Coat protein; Glycoprotein; Polyprotein; Signal; Transmembrane.
FT SIGNAL 1 20 Potential.
FT CHAIN 21 346 Outer membrane protein GP70 (Potential).
FT CHAIN 347 563 Transmembrane protein p20E (Potential).
FT CARBOHYD 113 113 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 219 219 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 229 229 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 264 264 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 282 282 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 292 292 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 306 306 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 312 312 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 321 321 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 339 339 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 469 469 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 563 AA; 61879 MW; 9573137DC46208B7 CRC64;

Query Match 22.0%; Score 580; DB 1; Length 563;
Best Local Similarity 29.2%; Pred. No. 8.2e-39;
Matches 146; Conservative 98; Mismatches 190; Indels 66; Gaps 14;

QY 28 TTSSSPYQELXRLPGLNIDAPSYSLKSGNSTFTTAHTMPRCNYSAPLCMHANTHW 87
DB 80 TSPSGPLQE-----SYQS-----SVHS-----SCYTSYQCRSGNKYY 117
QY 88 -----TGMINPSCPGGLGATVCTWYFTHTSMDGGGIGQAREK 127
DB 118 TATLLKTQTGGTSDVQLGVTNKLIOPCNGIKQSGICWSTTAPIHVSDDGGLDTRIK 177
QY 128 QVKEAISQLTRGHSSTPSYKGLVSLKHLTHRLVSLFNTLRLHEVSAQNPTNCW 187
DB 178 SVORKLEEIIKALYPELOYPHPLAIPKVRDLNLMVDAOTLNILNATYNLLMSNTSLVDDCW 237
QY 188 MCLPLHFRPYISIPVPEQMNFSSTEI--NTTSVLGVLPLVSNLBITHTSNLTCVKFNTID 245
DB 238 LCLKL--GPPTPLAIPNLLSYVTRSDNISCLIIPELL--VQPMQFSNSCL-FSPSYN 292
QY 246 TTS-----SQCIKRWTPTRIVCLPSGIFVCGTS--AYHCLNGSSSMCFSLFLVP 295
DB 293 STEEIDLGHVAFSNTS--ITNVTGPICAVNGSVFLCQNNMAYTLPNTWTGLCVLATLP 351
QY 296 PMTIYTEQD-----LYNHVVPKHNKRVPLPFVIRAGVLRLGTGIGSTTTSTQFYKL 350
DB 352 DIDIIPGDEVPPIADHFIYRP-KRAIQPIPLAGLIGITAAFTGATGLGVSVTQTKL 410
QY 351 SOEINGMEQVDSLVTLOQLNSLAVALQNRALDILLTAKRGCTCLFLGEERCYYVNO 410
DB 411 SNQLISDVQLLSSTIQDLQDVDSLAEVVLQNRGLDILLTAEQGGICLALQKCCFYVK 470
QY 411 SRIVTEKVEIRDIQCRABELQNTWGLLSQMPWLPFLGLAAILLLLLFGPCIFN 470
DB 471 SGIVRDKIKTQBELEERRRDKLASNPLWTGLQGLLPYLLPFLGLPFLTLLMLLTIGPCIFN 530
QY 471 LLVKFVSSRIEAVKLQNVLQ 490
DB 531 RLTAFLNDKLNIHAMVLQ 550

RESULT 20
Q9WP29 PRELIMINARY; PRT; 580 AA.
ID Q9WP29
AC Q9WP29;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN Name=env;
OS Simian retrovirus SRV-2.
OC Viruses; Retrovirdae; Retroviridae; Betaretrovirus.
OX NCBI_TaxID=39068;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D2/RHE/OR/V1;
RX MEDLINE=95191045; PubMed=7894914;
RA Marracci G.H., Kelley R.D., Pilcher K.Y., Crabtree L., Shiigi S.M., Avery N., Leo G., Webb M.C., Hallick L.M., Axthelm M.K., Machida C.A.; "Simian AIDS type D serogroup 2 retrovirus: isolation of an infectious molecular clone and sequence analyses of its envelope glycoprotein gene and 3' long terminal repeat.";
RL J. Virol. 69:2621-2628(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D2/RHE/OR/V1;
RX MEDLINE=99373443;
RA Marracci G.H., Avery N.A., Shiigi S.M., Couch G., Palmer H., Pilcher K.Y., Nichols H., Hallick L.M., Axthelm M.K., Machida C.A.; "Molecular cloning and cell-specific growth characterization of polymorphic variants of type D serogroup 2 simian retroviruses.";
RL Virology 261:43-58(1999).
DR EMBL; AF126468; AAD43261.1; --
DR GO; GO:0019028; C:Viral capsid; IEA.
DR GO; GO:0019031; C:Viral envelope; IEA.
DR GO; GO:0005198; F:Structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; TLV_coat; 1.
KW Envelope protein.
FT NON_TER 1 1
FT SEQUENCE 580 AA; 63668 MW; 84B03C8C45BE9457 CRC64;

Query Match 21.5%; Score 567.5; DB 2; Length 580;
Best Local Similarity 28.5%; Pred. No. 9.1e-38;
Matches 150; Conservative 99; Mismatches 203; Indels 75; Gaps 15;

QY 17 FALTAPPCCCTSSSPYQEFLLXRLPGLNIDAPSYSLK-----GN---STFTA 64
DB 54 YVITATPVYLATVSCSHSTAY-----QPS--DSLKRWCVSNPTLANGENIGNCPCTFKE 106
QY 65 HTHMPRCNYSATLCMHANTHWYG-----XMINPSCPGGLG 101
DB 107 SVH--SSCTTYTQECFFGNKTYTALASNAPTICTSNVPTVLGNTHNLSAGCTGNVG 164
QY 102 ATVCWTYFTHTSMDGGIGQAREKQVKEAISQLTRGHSSTPSYKGLVSLKHLTHRLTH 161
DB 165 QPICWNPAPVHISDGGGPDQARETAVQKRLLEEIHRSFLPELRYHPLALPKARGKEID 224
QY 162 TRLVSLFNTLRLHEVSAQNPTNCWMLPLHFRPYISIPVPEQMNFSSTINTTSLVVG 221
DB 225 AQTFNLLTATYSLLKNPNLANECWLCPSGNPVPFLAIPS-----NDSFLGSLNLSCPPI 280
QY 222 P-LVSNLEITHTSNLTCV--KFSNT-----IDTTSQCIRWTPTRIVCLPSGIGFF 270
DB 281 PLLVQPLEFINLINASCLISPPQNNSFGVDVGLVEFTNCSTLINI---SHSLCAPSSVF 337
QY 271 VCGTS--AYHCLNGSSSMCFSLPFLVPPMTIYTEQDLYNHVVPKP-----HNKR--VPIL 321
DB 338 VCGNNKAYTLPNTWTGTCVLATLLPDIDIVPG---DAPVPVPAIDHVLHARRAVQFI 393
QY 322 PFVIRAGVLRLGTGIGSTTTSTQFYKLQSOEINGMEQVDSLVTLOQLNSLAVALQ 381
DB 394 PLLVGLGITAVSTGTAGLGYISITQYTRLSRGLSDISVQAISSITQDLQDVDSLAELV 453
QY 382 NRRALDILLTAKRGCTCLFLGEERCYYVNOISRIVTEKVEIRDIQCRABELQNTWGLL 441
DB 454 NRRGLDILLTAEQGGICLALQKCCFYANKSGIVRDKIKRLQEDLEKRRKRIIDNPWTGL 513

QY 442 SQMPPWLPFLGPAALILLGLFGPCIFNLLVVFSSRIEAVKLOMW 488
 Db 514 HGLLPYLLPLGLLCLLLITTFGLIFNKIIAFVKQOMDAIQKPI 560

RESULT 21
 ENV_SRV2
 ID ENV_SRV2 STANDARD; PRT; 574 AA.
 AC PS1515;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Env polyprotein (Coat polyprotein) [Contains: Coat protein GP70; Coat protein GP20].
 DE protein GP20].
 GN Name=ENV;
 OS Simian retrovirus SRV-2.
 OC Viruses; Retroid viruses; Retroviridae; Betaretrovirus.
 OX NCBI_TaxID=39068;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87151131; PubMed=2435057;
 RA Thayer R.M., Power M.D., Bryant M.L., Gardner M.B., Barr P.J.,
 RA Luciw P.A.;
 RT "Sequence relationships of type D retroviruses which cause simian
 RT acquired immunodeficiency syndrome.";
 RL Virology 157:317-329(1987).
 CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR ENBL; M16605; AAA47563.1; ALT_INIT.
 DR HSSP; P03385; IMOF.
 DR InterPro; IPR002050; Env polyprotein.
 DR Pfam; PF00429; TLV_coat; 1.
 KW Coat protein; Glycoprotein; Polyprotein; Transmembrane.
 FT CHAIN 1 382
 FT CHAIN 383 574
 FT CHAIN 386 406
 FT TRANSMEM 386 574
 FT TRANSMEM 515 535
 FT CARBOHYD 117 117
 FT CARBOHYD 233 233
 FT CARBOHYD 260 260
 FT CARBOHYD 267 267
 FT CARBOHYD 288 288
 FT CARBOHYD 298 298
 FT CARBOHYD 312 312
 FT CARBOHYD 318 318
 FT CARBOHYD 327 327
 FT CARBOHYD 345 345
 FT CARBOHYD 475 475
 SQ SEQUENCE 574 AA; 62952 MW; 662B1E11A437E055 CRC64;

Query Match 21.4%; Score 562.5; DB 1; Length 574;
 Best Local Similarity 27.8%; Pred. No. 2.3e-37;
 Matches 151; Conservative 93; Mismatches 192; Indels 107; Gaps 15;

QY 24 PCCCT-----TSSSPYQBFLLKTRTRUPGNIDAPSVRLSK-----57
 Db 41 PDCAGYVNAAPTYVLAIVSCSSHTAYQP-----SDSLKWCVCVNPIANGENI 90
 QY 58 GN---STFTATHMPRCYNATLCMHANTHYWTG-----KM 91
 Db 91 GNCPCKTFKESVH--SSCYAYQCFPGNKTYTYTAILASNRAPIGTGTSNVTVLGNTHNL 148
 QY 92 INPSCPGGLGATVCWYTFHTTMSDGGGIGQAREKQVKEAISQLTRGHSTPPSYKGLVLSL 151
 Db 149 LSAGCTGNVGQPCICWNPKAPVPHISDGGGPDKAREIAVQKRLBEEIHKSLFPELYHPAL 208

QY 152 SKLHETLRTHTRLVSLFNTTLTRELHEVSAQNPTNCWMLPLHPRPYISIPVPSQWNNFST 211
 Db 209 PRARGKEKIDAQTFNLLTATYSLNKSNNPNELANECWCLP-SGNP---IPLAIPSNDSFL 264
 QY 212 EINTTSVLVGP-LVSNLEITHTSNLTCT-----VKFSNTIDTTSSQCIW 254
 Db 265 GSNLSCPIIPPLLVQPLEFPMNLINASCFFSPFONNSFDVDVGLVEFANCSSTLNI-----319
 QY 255 VTPPTTRIVCLPSGIFVFCGTS-AHCLNGSSSFCPLSFLVPMPTIYTEODLNVHVVPKP 313
 Db 320 ----SHSLCAPNSSVFCGNNKAYTYLPSNWTGTCVLATLLPDIDIVPG----DAPVPVP 371
 QY 314 -----HNKR--VPILPFVIRAGVRLGTGTSITTSQFYKLSQEIINGDMEQVTDLS 365
 Db 372 AIDHYLHRRARAVQFIPLLVGLGITTAVSTGTAGLGYSITQYTKLSKQLISDVQAISSTI 431
 QY 366 VTLQDQLNSLAALVQLNRRALDILLTAKRGTCICFLGEECCYYVQNSRIVTEKVEIRDRI 425
 Db 432 QDLQDQVDSLAELVQLNRRGLDILLTAEQGICLALQEKCCFYANKSGIVRDKIKRLOEDL 491
 QY 426 QCRAEELQNTERWGLLSQMMWVLPFLGPAALILLGLFGPCIFNLLVVFSSRIEAVKL 485
 Db 492 EKKRKEIIDNPFWTGLHGLPLLLPLGLFCLLLITFGFLIFNKIITFVKQIDAIOA 551
 QY 486 QMV 488
 Db 552 KPI 554

RESULT 22
 Q98654
 ID Q98654 PRELIMINARY; PRT; 564 AA.
 AC Q98654;
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Envelope protein.
 GN Name=env;
 OS RD114 retrovirus.
 OC Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses.
 OX NCBI_TaxID=11834;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96078985; PubMed=7494248;
 RA Cosses F.L., Takeuchi Y., Battini J.L., Weiss R.A., Collins M.K.L.;
 RT "High titer packaging cells producing recombinant retroviruses
 RT resistant to human serum";
 RL J. Virol. 69:7430-7436(1995).
 DR ENBL; X87829; CAA61093.1; -.
 DR HSSP; P03385; IMOF.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002050; Env polyprotein.
 DR Pfam; PF00429; TLV_coat; 1.
 KW Envelope protein.
 SQ SEQUENCE 564 AA; 62024 MW; 35F60EE7A51E91AC CRC64;

Query Match 21.2%; Score 559.5; DB 2; Length 564;
 Best Local Similarity 28.9%; Pred. No. 4e-37;
 Matches 134; Conservative 91; Mismatches 199; Indels 39; Gaps 8;

QY 63 TATHMPRCYNATLCMHANTHYWTGKMI-----NPS-----CPGGLGA 102
 Db 93 TFDQSMHSSCYTYRQCRINKTYTATLLKIRSGSINEVQIILQNPNQLQSCRSINQ 152
 QY 103 TVCWYTFHTTMSDGGGIGQAREKQVKEAISQLTRGHSTPPSYKGLVLSKLHETLRHT 162
 Db 153 PWCMSATAPIHISDGGGPDTRKVTQKRLQIHKAWTPELQYHPALPKVRRDLSLA 212
 QY 163 RLVSLENTLTRLHEVSAQNPTNCWMLPLHPRPYISIPVPSQWNNFSTINTSVLGP 222

Db 213 RTDILNTTFRLLQMSNFSIAQDCWLCCLKGTPTPLAIPTPSLTYSIADSLANASQIIP 272

Qy 223 --LVSNLEIHTHNTLVCVFSNT-----IDTSSQICIRWVTPTRIVCLPSGIFPFCV 273

Db 273 PLIVQPMQFNSCLSPFINDTEQIDLGAVTINCTSVANVSP---LCAALNGSVFLCG 329

Qy 274 TS-AYHCLNGSSBMCPLSLPVPMTIYTQD-----LYNHVVKPHKRVLPPLPVIRA 327

Db 330 NNMAVYLPQNWTRLCVQASLLPDIDINPGDEVPPIPAIDHYIHRP-KRAVQFIPILLAGL 388

Qy 328 GVLGRGTGGISITTSQFYFKLSQELNGMEQVTDLSVTLQDLSLAAVVLONBRALD 387

Db 389 GITAFTGTAGLVSVTVTKLSHQLSIDVQLSGTIQDLQDQVDSLAENVLQNRRLD 448

Qy 388 LLTAKRGTCFLGEERCYYVNSRIVTEKVIKIRDIRIQCRABELQWTERGWLISQWMPW 447

Db 449 LLTAEQGGICLAQEKCFYANKSGIVRNKIRTLQEBELQKRESLATNPLWTGLQGLFPY 508

Qy 448 VLPFLGPLAALILLFLGPGCIFNLLVKFVSSRIEAVKQLMWLQ 490

Db 509 LLPPLGLLTLTLITGPGCVFSRLMAFINDRNLNVHVMVLAQ 551

RESULT 23

ENV_AVISN STANDARD; PRT; 567 AA.

AC P31796;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 08-JUL-2004 (Rel. 44, Last annotation update)

DE Env polypotein (Coat polypotein) [Contains: Coat protein GP73; Coat protein GP22]

GN Names=ENV;

OS Avian spleen necrosis virus.

OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.

OX NCBI_TaxID=11899;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Clone PBH101;

RX MEDLINE=92219390; PubMed=1313915;

RA Kwaliramani V.N., Panganiban A.T., Emerman M.;

RT "Spleen necrosis virus, an avian immunosuppressive retrovirus, shares a receptor with the type D simian retroviruses.";

RL J. Virol. 66:3026-3031(1992).

CC -/- PFM: Specific enzymatic cleavages in vivo yield mature proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M87666; -; NOT_ANNOTATED_CDS.

DR PIR; A38212; VCFVAS.

DR HSSP; P03385; IMOF.

DR InterPro; IPR008965; Cellul bind.

DR InterPro; IPR002050; Env_polypotein.

DR Fram; PF00429; TLV_coat; 1.

KW Coat protein; Glycoprotein; Polyprotein; Transmembrane.

FT CHAIN 1 397

FT TRANSMEM 398 567

FT CARBOHYD 402 418

FT CARBOHYD 245 245

FT CARBOHYD 274 274

FT CARBOHYD 280 280

FT CARBOHYD 306 306

FT CARBOHYD 319 319

FT CARBOHYD 328 328

FT CARBOHYD 335 335

FT CARBOHYD 491 491

FT CARBOHYD 567 AA; 61596 MW; 204EA57C32159175 CRC64;

SQ SEQUENCE

Query Match 21.2%; Score 559.5; DB 1; Length 567;

Best Local Similarity 31.9%; Pred. No. 4e-37;

Matches 153; Conservative 75; Mismatches 158; Indels 93; Gaps 19;

Qy 66 THMPRCYNASATLCMEANTHYWT-----GKMNPSCPGGLGATVCTYFT 110

Db 121 TQMHSTCYKTOBCTLLGTYFTAILQKTKLGSYEDGNKLIQASCTGTGVKFCVDPVA 180

Qy 111 HTSMSGGGIQOAREKOVKEALSOLTRGHSTPS-PYKGLVLSKLHETLRTHTPLVSLFN 169

Db 181 PVTVDGGFTDMIRESVRLEEIIR-HSVPSVQYHPLALPR-----SRGVDLDP 231

Qy 170 TT---LTRLHEV-SAQNPP---TNCMCLPLHFRPYISIFVPEQWN-----NFSTEI--- 213

Db 232 QTSIDILEATHQVLNATNPXLAENCWLCMTL--GTPIPAAIPTNGVNLTDGNCSLSLPFGC 289

Qy 214 -----NTTSVLVGPLVSNLEIHTSNTLCVKSFTIDTSSQICIRWVTP 257

Db 290 NPPGSDIVSCYAGEADNRTGIPVG-----YVHFTNCTSIQ---EVTNETSQ---MGN 335

Qy 258 PTRIVCLPSGIFPFCVCTS-AYHCLNGSSBMCPLSLPVPMTIYTQDLYNHVVPKP--- 313

Db 336 LTRL-CPPGHVFCVGNMAYTALPNKWIGLCILASIVDPSISIGEE-----PIPLPSIE 390

Qy 314 -----HNKRVPLPFIIVIRAGVLGRL---GTGIGSITTSQFYFKLSQELNGMEQVTDLS 365

Db 391 YTARRHKRAVQFIPLLVGLGIGSATLAGGTGLG---VSVHTYHKLQNLIQEDVQALSGTI 447

Qy 366 VTLQDQLNSLAAVLQNRALDILLTAKRGTCFLGEERCYYVNSRIVTEKVKELRDRI 425

Db 448 NDLQDQIDSLAEVVLQNRRLDILLTAEQGGICLALQEKCCFYANKSGIIVRDKIRKQEDL 507

Qy 426 QCRABELQWTERGWLISQWMPVLPFLGPLAALILLFLGPGCIFNLLVKFVSSRIEAVK 484

Db 508 IERKRALYDNPLWSGLNGFLPILLPLGLFLGLILFLTGLPCIMKTLTRIHHDKIQAVK 566

RESULT 24

ENV_SRV1 STANDARD; PRT; 587 AA.

AC P04027;

DT 23-OCT-1986 (Rel. 02, Created)

DT 23-OCT-1986 (Rel. 02, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Env polypotein precursor (Coat polypotein) [Contains: Coat protein GP70; Coat protein GP20].

GN Name=ENV;

OS Simian retrovirus SRV-1.

OC Viruses; Retroid viruses; Retroviridae; Betaretrovirus.

OX NCBI_TaxID=11942;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=86151668; PubMed=3006247;

RA Power M.D., Marx P.A., Bryant M.L., Gardner M.B., Barr P.J.,

RA Luciw P.A.;

RT "Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus.";

RL Science 231:1567-1572(1986).

CC -/- PFM: Specific enzymatic cleavages in vivo yield mature proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M11841; AAA47733.1; -.

DR HSSP; P03385; IMOF.

DR InterPro; IPR002050; Env_polypotein.

DR Fram; PF00429; TLV_coat; 1.


```
Db      508 PFWTGLHGLLPYLLPILRLPILCLLLITFGPLFNKIIAFVKQMDAIQAKPI 560
RESULT 26
ENV_SRV26
ID ENV_SRV26 STANDARD; PRT; 574 AA.
AC P51520, 1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Env polyprotein (Coat polyprotein) [Contains: Coat protein GP70; Coat
DE protein GP20].
GN Name=ENV;
OS Simian retrovirus SRV-2 (isolate 2R-18B1).
OC Viruses; Retrovirdae; Retroviridae; Betaretrovirus.
OX NCBI_TaxID=73490;
RN [1]
RP MEDLINE=95191045; PubMed=7884914;
RA Maracci G.H., Kelley R.D., Pilcher K.Y., Crabtree L., Shiigi S.M.,
RA Avery N., Leo G., Webb M.C., Hallick L.M., Axthelm M.K., Machida A.;
RT "Simian AIDS type D serogroup 2 retrovirus: isolation of an infectious
RT molecular clone and sequence analyses of its envelope glycoprotein
RT gene and 3' long terminal repeat.";
RL J. Virol. 69:2621-2628(1995).
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; L38695; AAA68002.1; -.
DR HSP; P03385; IMOF.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; TIV_coat; 1.
KW Coat protein; Glycoprotein; Polyprotein; Transmembrane.
FT CHAIN 1 382 Coat protein GP70.
FT TRANSEM 383 574 Coat protein GP20.
FT TRANSSEM 386 406 Potential.
FT TRANSSEM 515 535 Potential.
FT CARBOHYD 117 117 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 233 233 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 260 260 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 267 267 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 288 288 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 298 298 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 312 312 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 318 318 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 327 327 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 345 345 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 475 475 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 574 AA; 63014 MW; 484EAF73ECABDD06 CRC64;
Query Match 21.0%; Score 553.5; DB 1; Length 574;
Best Local Similarity 28.0%; Pred. No. 1.3e-36;
Matches 149; Conservative 96; Mismatches 201; Indels 87; Gaps 14;
QY 17 FAUTAPPCCCTSSSPYQFLXRLPGLNIDAPSVLSK-----GN---STFTA 64
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
48 YVITAPTPLYATVSCSHTAY-----QPS--DSLKRCVSNPTLANGENIGNCPCQTFKE 100
QY 65 HTHMPCVNSATLCHMANTHYWG-----KMINSPCGLG 101
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
101 SVH--SSCYTYQECFGNKTYYTAILASNRAPFTGTSNPTVLGNTHLSAGCTGVG 158
QY 102 ATVCWTYFTHTSDGGIGQAREKQVKAISQLTRGHSPTSPYKGLVLSKLRLTH 161
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
159 QHICWNPAPVHISDGGPQDKAREITAVQKLEIHSRLPPELRYHPLALPKARGKEID 218
```

```
QY 162 TRLVSLNTTLRLHEVSAQNPTNCWMLPLHPRPYISIPVPEQWNNFSTEINTTSLVVG 221
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
219 AQTFNLLTATYSLINKSNPLANECLPFGNPNVPLAIPS-----NDSFLGSLSCPPIIP 274
QY 222 P-LVSNLEITHSTNLTC-----VKFSNTIDTSSQICIRWVTPPTTRIVCL 264
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
275 PLLVQPLEFINLINASCLYSPSQNSFDVDGLVFEINCSSTLNI-----SHSLCA 325
QY 265 PSGIFFVCGTS-AYHCLINGSSESMCLFLVLPMTIYTEQDLVNHVVPKP-----HNKR 317
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
326 PNSSVFVCGNKKAYTYLPTNWTGTCVLATLLPDIDIVPG---DAPVPAIDHYLHRR 381
QY 318 --VPILPFVIRAGVLRLGTGSGITTSQFYFKLSQEBNGMEQVTSLSVTLQDQNSL 375
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
382 RAVQFIPLLVGLGITTAVTSTGTLGYSITQTKLSQLISDVQAISITQDLQDQVDSL 441
QY 376 AAVLQNRALDILLTAKRGTCFLGEERCYYVQNSRIETKVKIQRDIQCRABELQNT 435
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
442 AEVLQNRRLGLDFTAEQGGICLAQKCCFYANKSGIVRDKIKALQEDLEKRRKEIIDN 501
QY 436 ERWGLLSQMPWVLPFLGLPGLAAILLLFLGPCIFNLLVFKVSSRIEAVKLMV 488
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
502 PFWTGLHGLLPYLLPILRLPILCLLLITFGPLFNKIIAFVKQMDAIQAKPI 554
RESULT 27
ENV_MPMV STANDARD; PRT; 586 AA.
ID ENV_MPMV STANDARD; PRT; 586 AA.
AC P07575;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Env polyprotein precursor (Coat polyprotein) [Contains: Coat protein
DE GP70; Coat protein GP20].
GN Name=ENV;
OS Simian Mason-Pfizer virus (MPMV).
OC Viruses; Retrovirdae; Retroviridae; Betaretrovirus.
OX NCBI_TaxID=11855;
RN [1]
RP SEQUENCE FROM N.A. (CLONE 6A).
RX MEDLINE=86189951; PubMed=2421920;
RA Sonigo P., Barker C., Hunter E., Wain-Hobson S.;
RT "Nucleotide sequence of Mason-Pfizer monkey virus: an
RT immunosuppressive D-type retrovirus.";
RL Cell 45:375-385(1986).
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M12349; AAA47712.1; -.
DR FIR; D25839; VCLJMP.
DR HSP; P03385; IMOF.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; TIV_coat; 1.
KW Coat protein; Glycoprotein; Polyprotein; Signal.
FT SIGNAL 1 20
FT CHAIN 21 394 Coat protein GP70.
FT CHAIN 395 586 Coat protein GP20.
FT CARBOHYD 120 120 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 237 237 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 264 264 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 276 276 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 291 291 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 304 304 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 318 318 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 324 324 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 339 339 N-linked (GlcNAc...) (Potential).
```


RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
RN [4]
RC SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=20493374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RC SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Tozawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RC SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata H., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito K., Saitoh H., Sakai K., Sakai K., Sakazume N., Sano H.,
RA Tagaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK086309; BAC39647.1; -
DR HSSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Tlv_coat; 1.
KW Envelope protein.
SQ SEQUENCE 618 AA; 69513 MW; 1D4B54588727FC31 CRC64;
Query Match 20.3%; Score 534.5; DB 2; Length 618;
Best Local Similarity 33.2%; Pred. No. 5.1e-35;
Matches 144; Conservative 71; Mismatches 134; Indels 85; Gaps 18;
QY 98 GGLGATVCMYFTHTSMDSGGGIGQAGREKQVKEAISQLTRGHSTPSPYKGG---LVLSK- 153
DB 183 GNLSASLC-NYTMHISPTSHENLDLSLTHTFQAMKRPDAKWNPLRSGPSPSLIFSKP 241
QY 154 -----LHETLRT-----HTRLVSLFNNTL-----TRLHEVS-AQNPFNCWML- 190
DB 242 AYYECPPTDIKCHTSPTATPMWCHCPQAPFGTCYNLTLPEDNTPHVTMSVNPHTFKVLQ 301
QY 191 -----PL-HFRYISIPVPEQNNSTNTSVLVGLPVSLELTHTSNLTCKVFSNT 243
DB 302 GHRDPYLSHYQPLTGAUSGGQYSVWNEIT-----VQENWDI--TSNI----FSHL 347
QY 244 IDTTSQCIRWVTPPTTRIVCLPSPGIFVCGTSAHYCHNGSSSMCFSLFVLPMTIYTEQ 303
DB 348 LSFSYAFCLN-----SSGVFLCGTSTYICLPANWSGVCTLVQYPIELLPN- 395

304 DLNVHVPKP-----HNKRV-ILPFIKRVAGLRLGTGIGSTTTTQFYKL 350
396 ---NQVPVPLFASVSSDLVRPKRSPHLPFPLAGLIGISSALGTIAGLATSTLYPQQL 452
351 SQEINGMEQVTDLSVTLQDLSLAADVLRALDLITAKGGTCLFLGEBRCYVYVQ 410
453 SKVLSLSEIAASITTLQNDLSLQVVLNRRALDLITAEKGGTCLFLQEECCFYVYVQ 512
411 SRVTVEKVEIRDRIQCRABEL-ONTERRGL---LSQMPWVLPFLGPIAALTLILLFGP 466
513 SGIVRDAARKLQE---RASELQCHSDSWGQWPDGLGRWLPDLTPFLGPLLFLFLITFGS 568
467 CIPNLLVKFVSSRI 480
569 CLLNCLTRFVSQRL 582

RESULT 30
Q9GLF7 PRELIMINARY; PRT; 620 AA.
AC O9GLF7;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Envelope protein.
GN Name=env;
OS Trichosurus vulpecula (Brush-tailed possum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.
OX NCBI_TaxID=9337;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21105294; PubMed=11160757;
RA Baillie G.J., Wilkins R.J.;
RT "Endogenous type D retrovirus in a marsupial, the common brushtail
RT possum (Trichosurus vulpecula).";
RL J. Virol. 75:2499-2507(2001).
DR EMBL; AF284693; AAG28161.1; -
DR HSSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Tlv_coat; 1.
KW Envelope protein.
SQ SEQUENCE 620 AA; 67973 MW; 12E1FF147B805E7F CRC64;
Query Match 19.9%; Score 524.5; DB 2; Length 620;
Best Local Similarity 28.1%; Pred. No. 3.4e-34;
Matches 134; Conservative 77; Mismatches 169; Indels 97; Gaps 12;
QY 72 CYNATLCMHANTHY-----W-----TGKMINPSCPG-GLGATVCW 106
DB 110 CYSSTQCTGKSGVLTSTQRAYGGSTGDNQGTQISGPTNKYAGASCDKINIGKVCW 169
QY 107 TYFTHTSMDSGGGIGQAGREKQVKEAISQLTRGHSTPSPYKGLVLSK-----LHET 157
DB 170 SLHAPTHVSDGGGPTDQIREMEVKERVDEIRAMPVPLQYHPLALPRGRGVLDVQTYDI 229
QY 158 LATHRLVSLFNNTLTRLHEVSAQNPTNCWMLPLHFRPYISIPVPEQNNSTNTTS 217
DB 230 LASTHKAALNTPDLAK-----DCWLCMTL-----GTPMPLALTLHDLSPATNC 273
QY 218 VLVGPL-----VSNLEITHTSNLTCKV----FSNTIDTTSQCIRW 254
DB 274 ALSPPFRVCPMPPLSAPCTEAPFRNSSYDNLVGHASFALCTSNHTFNVTVERT----- 326
QY 255 VTPPTTRIVCLPSPGIFVCGTSAHYCHNGSSSMCFSLFVLPMTIYTEQDLYNVHVPKP 313
DB 327 ----PHLLCPTGFAFVCGGNMAFLALPSNWTGLCVQASILPDINIISG----DQVPPLP 378
QY 314 H-----NKEVPILPFVIRAGVLRGLGTGIGSTTTTQFYKLQSOEINGMEQVTDLSL 365

Db 379 SIDYIAGRPKRAVAFIPLLVGLGVAGAMTTGSAGLGVAIHSAKLSNQLINDVQTLSGTI 438
Qy 366 VTLODQLNSLAADVVLQNRRLDILLTAKRGGTCLFLGEERCYYVNOQRIVTEKVKKEIRDRI 425
Db 439 HDLQDQIDSLAEVVVLQNRRLDILLTAEQGGTCLALQEKCCFYANKSGMVRDKIKLQOEEL 498
Qy 426 QCRAEELQNTERWGLLSQWMPWVLPELGPLAALTLILLFGPCIFNLVLKVFVSSRIEA 482
Db 499 VQRRKELNNPLNGLHGLLPYLLPLGLPLVGLLLLSFGFWVFNRTTTFVKSQVDS 555

Search completed: January 22, 2005, 22:11:46
Job time : 141 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2005, 10:09:14 ; Search time 5574.86 Seconds
(without alignments)
11273.477 Million cell updates/sec

Title: US-09-319-156B-12
Perfect score: 1329
Sequence: 1 tcaaaatcgaagagcttag.....gcaagatccccagtaaca 1329

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1326	99.8	1329	6 AR344389	Sequence
2	1326	99.8	1329	6 AX001030	Sequence
3	1326	99.8	1329	6 BD136199	Retrovira
4	1316.4	99.1	163803	9 AC093531	AC093531 Homo sapi
5	1076.8	81.0	10122	9 AY101590	Pongo pyg
6	1067	80.3	148711	9 AC133134	AC133134 Homo sapi
7	1064.8	80.1	10124	9 AY101591	Pongo pyg
8	1036.8	78.0	2694	9 AF520478	Homo sapi
9	1036.8	78.0	2694	9 AF520489S2	AF520489 Homo sapi
10	1036.8	78.0	2694	9 AF520505S2	AF520506 Homo sapi
11	1036.8	78.0	2694	9 AF520529S2	AF520530 Homo sapi
12	1035.2	77.9	2694	9 AF520483S2	AF520488 Homo sapi
13	1035.2	77.9	2694	9 AF520487S2	AF520494 Homo sapi
14	1035.2	77.9	2694	9 AF520493S2	AF520496 Homo sapi
15	1035.2	77.9	2694	9 AF520495S2	AF520498 Homo sapi
16	1035.2	77.9	2694	9 AF520497S2	AF520504 Homo sapi
17	1035.2	77.9	2694	9 AF520503S2	AF520510 Homo sapi
18	1035.2	77.9	2694	9 AF520509S2	AF520514 Homo sapi
19	1035.2	77.9	2694	9 AF520513S2	

20	1035.2	77.9	2694	9 AF520515S2	AF520516 Homo sapi
21	1035.2	77.9	2694	9 AF520517S2	AF520518 Homo sapi
22	1035.2	77.9	2694	9 AF520519S2	AF520520 Homo sapi
23	1035.2	77.9	2694	9 AF520521S2	AF520522 Homo sapi
24	1035.2	77.9	2694	9 AF520525S2	AF520526 Homo sapi
25	1035.2	77.9	2694	9 AF520533S2	AF520534 Homo sapi
26	1035.2	77.9	2694	9 AF520535S2	AF520536 Homo sapi
27	1035.2	77.9	2694	9 AF520537S2	AF520538 Homo sapi
28	1035.2	77.9	2694	9 AF520549S2	AF520550 Homo sapi
29	1035.2	77.9	2694	9 AF520553S2	AF520554 Homo sapi
30	1035.2	77.9	2694	9 AF520561S2	AF520562 Homo sapi
31	1035.2	77.9	2694	9 AF520563S2	AF520564 Homo sapi
32	1035.2	77.9	56093	6 AX329572	AX329572 Sequence
33	1035.2	77.9	56093	9 HSAC000064	AC000064 Human BAC
34	1035.2	77.9	149194	9 AC007566	AC007566 Homo sapi
35	1033.6	77.8	2694	9 AF520485S2	AF520486 Homo sapi
36	1033.6	77.8	2694	9 AF520491S2	AF520492 Homo sapi
37	1033.6	77.8	2694	9 AF520499S2	AF520500 Homo sapi
38	1033.6	77.8	2694	9 AF520501S2	AF520502 Homo sapi
39	1033.6	77.8	2694	9 AF520507S2	AF520508 Homo sapi
40	1033.6	77.8	2694	9 AF520511S2	AF520512 Homo sapi
41	1033.6	77.8	2694	9 AF520523S2	AF520524 Homo sapi
42	1033.6	77.8	2694	9 AF520527S2	AF520528 Homo sapi
43	1033.6	77.8	2694	9 AF520531S2	AF520532 Homo sapi
44	1033.6	77.8	2694	9 AF520539S2	AF520540 Homo sapi
45	1033.6	77.8	2694	9 AF520541S2	AF520542 Homo sapi

ALIGNMENTS

RESULT 1
AR344389
LOCUS AR344389 1329 bp mRNA linear PAT 17-AUG-2003
DEFINITION Sequence 108 from patent US 6582703.
ACCESSION AR344389
VERSION AR344389.1 GI:33740330
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1329)
AUTHORS Perron,H., Beseme,F., Bedin,F., Paranhos-Baccala,G., Komurian-Pradel,F., Jolivet-Reynaud,C. and Mandrand,B.
TITLE Isolated nucleotide sequences associated with multiple sclerosis or rheumatoid arthritis and a process of detecting
JOURNAL Patent: US 6582703-A 108 24-JUN-2003;
FEATURES
source Location/Qualifiers
1..1329
/organism="unknown"
/mol_type="mRNA"

ORIGIN

Query Match 99.8%; Score 1326; DB 6; Length 1329;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAGAGGGGAACTGTTTATTTT 60
Db 1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAGAGGGGAACTGTTTATTTT 60
Qy 61 AGGGGAAGAAATGCTGTTAGTATGTTAATCAATCTGGAATCATTTCTGAGAAGTAAAGA 120
Db 61 AGGGGAAGAAATGCTGTTAGTATGTTAATCAATCTGGAATCATTTCTGAGAAGTAAAGA 120
Qy 121 AATTGAGATCGAATATAATGTAGACAGAGGACCTTCAAAACATCTGACCTGGGGCCT 180
Db 121 AATTGAGATCGAATATAATGTAGACAGAGGACCTTCAAAACATCTGACCTGGGGCCT 180
Qy 181 CCTCAGCAATGATGCTCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTAATATT 240
Db 181 CCTCAGCAATGATGCTCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTAATATT 240

QY 241 TTATCTCTCTTTGGACCCCTGTATCTTCAACTTCCTCTTTAAAGTTTGTCTCTTCCAGAAAT 300
DB 241 TTTACTCTCTTTGGACCCCTGTATCTTCAACTTCCTCTTTAAAGTTTGTCTCTTCCAGAAAT 300
QY 301 TGAAGCTGTAAGACTACAAATAGTCTTCAAAATGGAAACCCAGATGCAATGCACTAA 360
DB 301 TGAAGCTGTAAGACTACAAATAGTCTTCAAAATGGAAACCCAGATGCAATGCACTAA 360
QY 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGTATGTTAATGACATTGA 420
DB 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGTATGTTAATGACATTGA 420
QY 421 AGTCACCCCTCCGAGGAAATCTCAACTGCACAAACCCCTACTACACTCCAAATTCAGTAGG 480
DB 421 AGTCACCCCTCCGAGGAAATCTCAACTGCACAAACCCCTACTACACTCCAAATTCAGTAGG 480
QY 481 AAGCAGTTAGACAGTGTTCAGCAACCTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 540
DB 481 AAGCAGTTAGACAGTGTTCAGCAACCTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 540
QY 541 GGGTGGACTGAGACAGGACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCNAAAGCCT 600
DB 541 GGGTGGACTGAGACAGGACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCNAAAGCCT 600
QY 601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
DB 601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
QY 661 ACCAATCAGAGAGTCACTAAATGCTAATCAGGCAAAAACAGGAGTAAAGCAATAGCC 720
DB 661 ACCAATCAGAGAGTCACTAAATGCTAATCAGGCAAAAACAGGAGTAAAGCAATAGCC 720
QY 721 AATCATCTATTGCTGAGACACAGCGGGAAGGACAGGATTTGGGATATAAACTCAGGCA 780
DB 721 AATCATCTATTGCTGAGACACAGCGGGAAGGACAGGATTTGGGATATAAACTCAGGCA 780
QY 781 TTCAAGCAGCAACAGCAACCCCTTTGGGTCCCTCCCACTGTATGGGAGCTCTGTTT 840
DB 781 TTCAAGCAGCAACAGCAACCCCTTTGGGTCCCTCCCACTGTATGGGAGCTCTGTTT 840
QY 841 CACTCTATTCACTCTATTAAATCATGCACTGCACTCTCTCTGTCGGTGTATTTATGG 900
DB 841 CACTCTATTCACTCTATTAAATCATGCACTGCACTCTCTCTGTCGGTGTATTTATGG 900
QY 901 CTCAAGCTGAGCTTTGTCGCCATCCACCACTGCTTTGGCCACCGTACAGACCCGCT 960
DB 901 CTCAAGCTGAGCTTTGTCGCCATCCACCACTGCTTTGGCCACCGTACAGACCCGCT 960
QY 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCA CTGTGCTCTGATCCAGCAGGT 1020
DB 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCA CTGTGCTCTGATCCAGCAGGT 1020
QY 1021 ACCCATGCGACCTCCGATCAGGCTAAAGGCTTGGCAATGTTCTCTGATGGCTAAGTGC 1080
DB 1021 ACCCATGCGACCTCCGATCAGGCTAAAGGCTTGGCAATGTTCTCTGATGGCTAAGTGC 1080
QY 1081 TGGGTTTGCTTAATAGAACTGAACACTGCTCACTGGGTTCCATGTTCTCTTCCATGAC 1140
DB 1081 TGGGTTTGCTTAATAGAACTGAACACTGCTCACTGGGTTCCATGTTCTCTTCCATGAC 1140
QY 1141 CCAAGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
DB 1141 CCAAGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGGTA 1200
QY 1201 TCTGTAGGCAAGAACCCAGGTACAGAAAGTGGGCTTGGCAACATTTGGGAAGTGG 1260
DB 1201 TCTGTAGGCAAGAACCCAGGTACAGAAAGTGGGCTTGGCAACATTTGGGAAGTGG 1260
QY 1261 CCCACTGCAATTTTGGTAGGGGCCCAACCATCTTTGGGAGCTGTGGGAGCAAGGATCCC 1320
DB 1261 CCCACTGCAATTTTGGTAGGGGCCCAACCATCTTTGGGAGCTGTGGGAGCAAGGATCCC 1320
QY 1321 CCAGTAACA 1329

DB 1321 CCAGTAACA 1329
RESULT 2
AX001030 1329 bp DNA linear PAT 10-MAR-2000
LOCUS Sequence 12 from Patent WO9902666.
DEFINITION AX001030
ACCESSION AX001030
VERSION AX001030.1 GI:7241266
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 1329)
AUTHORS Ott C. and Bedin, F.
TITLE RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR
ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR
DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES
JOURNAL Patent: WO 9902666-A 12 21-JAN-1999;
BIO MERIEUX (FR); OTT CATHERINE (FR)
FEATURES
source Location/Qualifiers
1..1329 /organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
ORIGIN
Query Match 99.8%; Score 1326; DB 6; Length 1329;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCAAAATCGAAGAGCTTTAGACTTTGCTAAACCGCAAAAGAGGGGAAACCTGTTTATTTT 60
DB 1 TCAAAATCGAAGAGCTTTAGACTTTGCTAAACCGCAAAAGAGGGGAAACCTGTTTATTTT 60
QY 61 AGGGGAAGAAATGCTGTTAGTATGTTTAAATCAATCTGGAATCATTTACGAGAAAGTTAAAGA 120
DB 61 AGGGGAAGAAATGCTGTTAGTATGTTTAAATCAATCTGGAATCATTTACGAGAAAGTTAAAGA 120
QY 121 AATTTGAGATCGAATATAATGTAGACAGAGACCTTCAAAACACTGCAACCTCGGGGCT 180
DB 121 AATTTGAGATCGAATATAATGTAGACAGAGACCTTCAAAACACTGCAACCTCGGGGCT 180
QY 181 CTTACGCCAATGATGCTCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
DB 181 CTTACGCCAATGATGCTCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
QY 241 TTTACTCTCTTTGGACCCCTGTATCTTCAACTTCCTTTGTTAAAGTTTGTCTCTTCCAGAAAT 300
DB 241 TTTACTCTCTTTGGACCCCTGTATCTTCAACTTCCTTTGTTAAAGTTTGTCTCTTCCAGAAAT 300
QY 301 TGAAGCTGTAAGACTACAAATAGTCTTCAAAATGGAAACCCAGATGCAATGCACTAA 360
DB 301 TGAAGCTGTAAGACTACAAATAGTCTTCAAAATGGAAACCCAGATGCAATGCACTAA 360
QY 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGTATGTTAATGACATTGA 420
DB 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGTATGTTAATGACATTGA 420
QY 421 AGTCACCCCTCCGAGGAAATCTCAACTGCACAAACCCCTACTACACTCCAAATTCAGTAGG 480
DB 421 AGTCACCCCTCCGAGGAAATCTCAACTGCACAAACCCCTACTACACTCCAAATTCAGTAGG 480
QY 481 AAGCAGTTAGACAGTGTTCAGCAACCTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 540
DB 481 AAGCAGTTAGACAGTGTTCAGCAACCTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 540
QY 541 GGGTGGACTGAGACAGGACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCNAAAGCCT 600
DB 541 GGGTGGACTGAGACAGGACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCNAAAGCCT 600
QY 601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660

Db 601 ANCTGGAGGTGACCGCATCTTAAACATGGGCTTGCACTTAGCTCACCCG 660
Qy 661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAAAACAGGAGTAAAGCAATAGCC 720
Db 661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAAAACAGGAGTAAAGCAATAGCC 720
Qy 721 AATCATCTATTGCTGAGAGCAGCGGGAAGCAAGGATTTGGGATATATAACTCAGGCA 780
Db 721 AATCATCTATTGCTGAGAGCAGCGGGAAGCAAGGATTTGGGATATATAACTCAGGCA 780
Qy 781 TTCAAGCCAGCAACAGCAACCCCTTTGGGTGCCCTCCCATTTGATGGGAGCTCTGTTTT 840
Db 781 TTCAAGCCAGCAACAGCAACCCCTTTGGGTGCCCTCCCATTTGATGGGAGCTCTGTTTT 840
Qy 841 CACTCTATTCTACTCTATTAATCATGCAACTGCACTCTTTCCGTCGTTTTTATGG 900
Db 841 CACTCTATTCTACTCTATTAATCATGCAACTGCACTCTTTCCGTCGTTTTTATGG 900
Qy 901 CTCAGCTGAGCTTTGTTCCCATCCACACACTGCTGTTGGCCACCGTCAACACCGCT 960
Db 901 CTCAGCTGAGCTTTGTTCCCATCCACACACTGCTGTTGGCCACCGTCAACACCGCT 960
Qy 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCACACTGCTCCTGATCCAGCGAGST 1020
Db 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCACACTGCTCCTGATCCAGCGAGST 1020
Qy 1021 ACCATTTGCCACTCCCGATCAGGCTAAAGGCTTTGCCATTTGCTGATGGCTAAAGTGCC 1080
Db 1021 ACCATTTGCCACTCCCGATCAGGCTAAAGGCTTTGCCATTTGCTGATGGCTAAAGTGCC 1080
Qy 1081 TGGGTTTGCTTAATAGAACTGAACACTGTGTCTCTGGTTCCATGTTCTCTCCATGAC 1140
Db 1081 TGGGTTTGCTTAATAGAACTGAACACTGTGTCTCTGGTTCCATGTTCTCTCCATGAC 1140
Qy 1141 CCAGGCTTCTAATAGAGCTATACACTCACCGCATGGCCCAAGATTTCCATTTCTTGTA 1200
Db 1141 CCAGGCTTCTAATAGAGCTATACACTCACCGCATGGCCCAAGATTTCCATTTCTTGTA 1200
Qy 1201 TCTGTGAGGCAAGAACCCAGGTCAGAGAAAGTGAGGCTTGCCACCATTTGGGAAGTGG 1260
Db 1201 TCTGTGAGGCAAGAACCCAGGTCAGAGAAAGTGAGGCTTGCCACCATTTGGGAAGTGG 1260
Qy 1261 CCACATGCCATTTTGGTGGGCCCCACCACTCTTTGGGAGCTGTGGGAGCAAGATGCC 1320
Db 1261 CCACATGCCATTTTGGTGGGCCCCACCACTCTTTGGGAGCTGTGGGAGCAAGATGCC 1320
Qy 1321 CCAGTAACA 1329
Db 1321 CCAGTAACA 1329

RESULT 3
BD136199
LOCUS
DEFINITION
BD136199 1329 bp DNA linear PAT 18-SEP-2002
Retroviral nucleic material and nucleotide fragments, in
particular, associated with multiple sclerosis and/or rheumatoid
arthritis, for diagnostic, prophylactic and therapeutic uses.
ACCESSION
BD136199
VERSION
BD136199.1 GI:23231144
KEYWORDS
JP 2002509437-A/9.
SOURCE
unidentified
ORGANISM
unclassified.
REFERENCE
1 (bases 1 to 1329)
AUTHORS
Baccala,G.P., Pradel,F.K., Bedin,F., Sodoyer,M., Ott,C., Mallet,F.,
Perron,H. and Mandrand,B.
TITLE
Retroviral nucleic material and nucleotide fragments, in
particular, associated with multiple sclerosis and/or rheumatoid
arthritis, for diagnostic, prophylactic and therapeutic uses
JOURNAL
Patent: JP 2002509437-A 9 26-MAR-2002;
BIO MERIEUX
COMMENT
OS Unidentified

PN JP 2002509437-A/9
PD 26-MAR-2002
PF 07-JUL-1998 JP 1999508255
PR 07-JUL-1997 FR 97/08816
PI GLAUCIA PARANHOS BACCALA,FLORENCE KOMURIAN PRADEL,FREDERIC PI
BEDIN,
PI MIREILLE SODOYER,CATHERINE OTT,FRANCOIS MALLET,HERVE PERRON,
PI BERNARD MANDRAND
PC C12N15/48,C12Q1/70,C07K14/15,A61K31/70
CC Strandedness: Single;
CC Topology: Linear;
CC Retroviral nucleic material and nucleotide fragments, in CC
particular, associated with multiple sclerosis and/or
rheumatoid arthritis,
CC for
CC diagnostic, prophylactic and therapeutic uses FH Key
Location/Qualifiers
FT source 1..1329 /organism='Unidentified'.
FT Location/Qualifiers
1..1329 /organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'
Query Match 99.8%; Score 1326; DB 6; Length 1329;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TCAAAATCGAAGAGCTTTAGACTGCTAACCGCCAAAGAGGGGGAACCTGTTATTTT 60
Db 1 TCAAAATCGAAGAGCTTTAGACTGCTAACCGCCAAAGAGGGGGAACCTGTTATTTT 60
Qy 61 AGGGGAAGAGTGTGTAGTATGTTAATCAATCTGGAATCATTAATCTGAGAAAGTAAAGA 120
Db 61 AGGGGAAGAGTGTGTAGTATGTTAATCAATCTGGAATCATTAATCTGAGAAAGTAAAGA 120
Qy 121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTTCAAAACACATGACCCCTGGGCCT 180
Db 121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTTCAAAACACATGACCCCTGGGCCT 180
Qy 181 CCTCAGCAATAGAGTCCCTGGACTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATT 240
Db 181 CCTCAGCAATAGAGTCCCTGGACTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATT 240
Qy 241 TTTTACTCTCTTTGGACCCCTGTATCTCAACTCTCTGTTAAAGTTCTCTTCCAGAT 300
Db 241 TTTTACTCTCTTTGGACCCCTGTATCTCAACTCTCTGTTAAAGTTCTCTTCCAGAT 300
Qy 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAACCCAGATGAGTCCATGACTAA 360
Db 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAACCCAGATGAGTCCATGACTAA 360
Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTGTATGATGATTAATGACATGA 420
Db 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTGTATGATGATTAATGACATGA 420
Qy 421 AGTCACCCCTCCCGAGGAATCTCAACTGCAACCCCTACTACACTCCCAATTCAGTAGG 480
Db 421 AGTCACCCCTCCCGAGGAATCTCAACTGCAACCCCTACTACACTCCCAATTCAGTAGG 480
Qy 481 AAGCAGTTAGAGCAGTTGTTCAGGCAACCTCCCAACAGTACTTTGGGTTTCTCTTTGAGA 540
Db 481 AAGCAGTTAGAGCAGTTGTTCAGGCAACCTCCCAACAGTACTTTGGGTTTCTCTTTGAGA 540
Qy 541 GGGTGGACTGAGAGCAGGACTAGCTGGAATTTCTAGGCTGACTAAGAAATCCCNAGCCT 600
Db 541 GGGTGGACTGAGAGCAGGACTAGCTGGAATTTCTAGGCTGACTAAGAAATCCCNAGCCT 600
Qy 601 ANCTGGAGGTGACCGCATCTTAAACATGGGCTTGCACTTAGCTCACCCG 660

```

Db 601 ANCTGGGAAGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTAGCTCACACCG 660
Qy 661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAAAACAGGAGTAAAGCAATAGCC 720
Db 661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAAAACAGGAGTAAAGCAATAGCC 720
Qy 721 AATCATCTATTGCTGAGAGACACAGCGGGAAGGAAGGATTTGGGATATATAACTCAGGCA 780
Db 721 AATCATCTATTGCTGAGAGACACAGCGGGAAGGAAGGATTTGGGATATATAACTCAGGCA 780
Qy 781 TTAAGCAGCAACAGCAACCCCTTTGGGTGCGCTCCCATTTGATGGGAGCTCTGTTTT 840
Db 781 TTAAGCAGCAACAGCAACCCCTTTGGGTGCGCTCCCATTTGATGGGAGCTCTGTTTT 840
Qy 841 CACTCTATTTCACTCTATTAATCATGCAACTGCACTCTTCTGGTCCGTTGTTTTATGG 900
Db 841 CACTCTATTTCACTCTATTAATCATGCAACTGCACTCTTCTGGTCCGTTGTTTTATGG 900
Qy 901 CTCAAGCTGAGCTTTGTTGCGCATCCACCACTGCTGTTTGCCACCGTCAAGACCGCT 960
Db 901 CTCAAGCTGAGCTTTGTTGCGCATCCACCACTGCTGTTTGCCACCGTCAAGACCGCT 960
Qy 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCACCTGCTCCTGATCCAGCGAGGT 1020
Db 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCACCTGCTCCTGATCCAGCGAGGT 1020
Qy 1021 ACCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCTGTCATGGCTAAAGTCC 1080
Db 1021 ACCATTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCTGTCATGGCTAAAGTCC 1080
Qy 1081 TGGGTTGTCCTAATAGAACTGAACACTGCTGCTCACTGGTTCCTTCCTCATGAC 1140
Db 1081 TGGGTTGTCCTAATAGAACTGAACACTGCTGCTCACTGGTTCCTTCCTCATGAC 1140
Qy 1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTTCCATTCTTGGA 1200
Db 1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTTCCATTCTTGGA 1200
Qy 1201 TCTGTGAGGCAAGAACCCAGGCTCAGAGAAAGTGGGCTTGCCACCATTTGGGAAGTGG 1260
Db 1201 TCTGTGAGGCAAGAACCCAGGCTCAGAGAAAGTGGGCTTGCCACCATTTGGGAAGTGG 1260
Qy 1261 CCACATGCCAATTTGGTAGGGGCCACCACTCTGGGAGCTGTGGGAGCAAGGATCCC 1320
Db 1261 CCACATGCCAATTTGGTAGGGGCCACCACTCTGGGAGCTGTGGGAGCAAGGATCCC 1320
Qy 1321 CCAGTAACA 1329
Db 1321 CCAGTAACA 1329

```

```

RESULT 4
AC093531/C 163803 bp DNA linear PRI 16-NOV-2001
LOCUS Homo sapiens chromosome 5 clone RP11-405L7, complete sequence.
DEFINITION AC093531
ACCESSION AC093531
VERSION AC093531.2 GI:16945981
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 163803)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
2 (bases 1 to 163803)
DOE Joint Genome Institute.
AUTHORS Direct Submission
TITLE Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 163803)
DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS

```

```

TITLE Direct Submission
JOURNAL Submitted (16-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Nov 16, 2001 this sequence version replaced gi:15383820.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.3.
FEATURES
source Location/Qualifiers
1.163803
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-405L7"
ORIGIN

```

```

Query Match 99.1%; Score 1316.4; DB 9; Length 163803;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1320; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
Qy 1 TCAAAATCGAAGAGCTTTAGACTTGTAAATCAATCTGGAATCATTTACTGAGAAAGTTAAAGA 60
Db 117604 TTAATAATTGAAGAGCTTTAGACTTGTAAATCAATCTGGAATCATTTACTGAGAAAGTTAAAGA 117545
Qy 61 AGGGGAAGAAATGCTGTTAGTATGTTAAATCAATCTGGAATCATTTACTGAGAAAGTTAAAGA 120
Db 117544 AGGGGAAGAAATGCTGTTAGTATGTTAAATCAATCTGGAATCATTTACTGAGAAAGTTAAAGA 117485
Qy 121 AATTTGAGATCGAATAATAATGTAGAGCAGAGGACCTTCAAAAACATGACCCCTGGGCGCT 180
Db 117484 AATTTGAGATCGAATAATAATGTAGAGCAGAGGACCTTCAAAAACATGACCCCTGGGCGCT 117425
Qy 181 CCTCAGCCAATGATGCCCTGGAGCTCTCCCTCTTTAGGACCTCTAGCAGCTATAATATT 240
Db 117424 CCTCAGCCAATGATGCCCTGGAGCTCTCCCTCTTTAGGACCTCTAGCAGCTATAATATT 117365
Qy 241 TTTACTCTCTTTTGGACCCCTGATCTTCAACTTCTTGTAAAGTTGTCTCTCCAGAAT 300
Db 117364 TTTACTCTCTTTTGGACCCCTGATCTTCAACTTCTTGTAAAGTTGTCTCTCCAGAAT 117305
Qy 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAAACCCAGATGCAATGATGATAA 360
Db 117304 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAAACCCAGATGCAATGATGATAA 117245
Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
Db 117244 AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGTTAATGACATTGA 117185
Qy 421 AGTCACCCCTCCCGAGGAAATCTCAACTGCAACACCCCTACTACACTCCAAATTCAGTAGG 480
Db 117184 AGTCACCCCTCCCGAGGAAATCTCAACTGCAACACCCCTACTACACTCCAAATTCAGTAGG 117125
Qy 481 AAGCAGTTAGAGCAGTTGTGACGCAACCTCCCAACAGTACTTTGGGTTTCTGTTGAGA 540
Db 117124 AAGCAGTTAGAGCAGTTGTGACGCAACCTCCCAACAGTACTTTGGGTTTCTGTTGAGA 117065
Qy 541 GGGTGGACTGAGACAGGACTAGCTGGATTTCTAGGCTGACTAAGAAATCCCNAGCCT 600
Db 117064 GGGTGGACTGAGACAGGACTAGCTGGATTTCTAGGCTGACTAAGAAATCCCNAGCCT 117005
Qy 601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTAGCTCACACCG 660
Db 117004 AGCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTAGCTCACACCG 116945
Qy 661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAAAACAGGAGTAAAGCAATAGCC 720
Db 116944 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAAAACAGGAGTAAAGCAATAGCC 116885
Qy 721 AATCATCTATTGCTGAGAGACACAGCGGGAAGGAAGGATTTGGGATATATAACTCAGGCA 780

```

```
Db 116884 AATCATCTATTGCTGAGAGACAGCGGGAAGGACAAAGGATTGGGATATAAACCAGGCA 116825
Qy 781 TTCAAGCAGCAACAGCAACCCCTTTGGGTCCCTCCATTTGATGGAGCTCTGTTTT 840
Db 116824 TTCAAGCAGCAACAGCAACCCCTTTGGGTCCCTCCATTTGATGGAGCTCTGTTTT 116765
Qy 841 CACTCTATTTTCACTCTATTAATCAATCATCACTGACCTCTCTGCTCGTGTGTTTTATGG 900
Db 116764 CACTCTATTTTCACTCTATTAATCAATCATCACTGACCTCTCTGCTCGTGTGTTTTATGG 116705
Qy 901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGGCCACCGTCCACAGCCGCT 960
Db 116704 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGGCCACCGTCCACAGCCGCT 116645
Qy 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCACCTGTGCTCTGATCCAGCGAGGT 1020
Db 116644 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCACCTGTGCTCTGATCCAGCGAGGT 116585
Qy 1021 ACCATTGCCACTCCCATCAGGCTAAGGCTTGCCATTTGCTGCTGATCCAGCGAGGT 1080
Db 116584 ACCATTGCCACTCCCATCAGGCTAAGGCTTGCCATTTGCTGCTGATCCAGCGAGGT 116525
Qy 1081 TGGTTTGTCTTAATGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTCCATGAC 1140
Db 116524 TGGTTTGTCTTAATGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTCCATGAC 116465
Qy 1141 CCACGGCTTCTAATAGAGCTATAACACTACCGCATCGCCCAAGATTCCATTCCTGGTA 1200
Db 116464 CCACGGCTTCTAATAGAGCTATAACACTACCGCATCGCCCAAGATTCCATTCCTGGTA 116405
Qy 1201 TCTGTAGGCCAAGAACCCAGGTCAGAAAGTGAGGCTTGCCACCATTTGGGAAGTGG 1260
Db 116404 TCTGTAGGCCAAGAACCCAGGTCAGAAAGTGAGGCTTGCCACCATTTGGGAAGTGG 116345
Qy 1261 CCACACTGCCATTTGGTAGGGCCACACCATCTTGGGAGCTGTGGGAGCAAGATCCC 1320
Db 116344 CCACACTGCCATTTGGTAGGGCCACACCATCTTGGGAGCTGTGGGAGCAAGATCCC 116285
Qy 1321 CCAGTAACA 1329
Db 116284 CCAGTAACA 116276

RESULT 5
LOCUS AV101590 10122 bp DNA linear PRI 11-FEB-2004
DEFINITION Pongo pygmaeus isolate 1 endogenous retrovirus ERV-W, ERVW1 locus,
allele A, complete sequence.
ACCESSION AV101590
VERSION AV101590.1 GI:37544415
KEYWORDS
SOURCE Pongo pygmaeus (orangutan)
ORGANISM Pongo pygmaeus
Sukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
1 (bases 1 to 10122)
Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
The endogenous retroviral locus ERVW1 is a bona fide gene involved
in hominoid placental physiology
Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
14757826
2 (bases 1 to 10122)
Mallet,F., Bouton,O. and Oriol,G.
Direct Submission
Submitted (06-MAY-2002) Retrovirology Department, UMR 2142
CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée
d'Italie, Lyon 69364 cedex 07, France
Location/Qualifiers
1..10122
/organism="Pongo pygmaeus"
/mol_type="genomic DNA"
/isolate="1"
```

```
/isolation_source="PBMC of orangutan 1"
/db_xref="taxon:9600"
/cell_type="PBMC"
/notes="isolated by PCR; ERVW1 locus, allele A
endogenous_virus: ERV-W"
1..780
join(247..811,2872..9669)
/notes="putative mRNA transcript 1; gag-pol pseudogene"
join(247..811,7440..9669)
/product="envelope glycoprotein"
/notes="putative mRNA transcript 2"
join(247..811,9104..9669)
/notes="putative mRNA transcript 3; no coding region"
7668..9284
/notes="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAM68169.1"
/db_xref="GI:37544416"
/translation="MALPHYHIFLFTVLLPSFTLTAPPCCMTSSSPYQBFLLPMHHP
GNIDAPYSFSGTPTPTAHTHMPNCINSATLCMRGNTHTYTGKMINSCPGGLGV
TVCMYFTHTGMSDGGGVQDEAREKHVKEVISQITQVHSSPYKGLDLKSLHETLRT
LTVGLVSLFNLTTLGLHEVSAQNTNCWMLPLAFRPYVSIIPVPEQMNFFTEINTTSV
HPLGLVSNLEIHTLSNLTKVKSNTNTNTNSQCIRVPTQVVCPLPSGIFPVGCTSA
YRCLNGSSSMCFSLFLVPMPTIYTBQDLNVYVSKPRKRPVILPFWMAAGVLGALG
TGIGITTSQFYKLSQELNGDMERVADSLVTLDQNLNSIAAVLQNRRLDLITAE
RGCTCLFGECCYYNKGIVTEKVKIRDIORRABELRNIGPWGLFSOMMEWILP
FLGPLAAIILLLLFGPCIFNLNVSSRIEAVKLQMEPKQSKTKIYHRPLDWPASP
RSDVNDIKGTTPPEIESTAQLLRPNAGSS"
9343..10122

LTR
ORIGIN
Query Match 81.0%; Score 1076.8; DB 9; Length 10122;
Best Local Similarity 89.2%; Pred. No. 0;
Matches 1186; Conservative 0; Mismatches 130; Indels 13; Gaps 2;
Qy 1 TCAAAATCGAAGACTTTAGACTGCTAACCGCCAAAGAGGGGGAACCTGTTATTTT 60
Db 8807 TCAAAATCGAAGACTTTAGACTGCTAACCGCCAAAGAGGGGGAACCTGTTATTTT 8866
Qy 61 AGGGGAAGATGCTGTTAGTATGTTAATCAATCTGGAATCATTAATCTGAGAAAGTTAAAGA 120
Db 8867 AGGGGAAGATGCTGTTATGTTAATCAATCGGAATTTGCTACTGAGAAAGTTAAAGA 8926
Qy 121 AATTTGAGATCGAATAATATGTAGAGCAGAGGACCTTCAAAACACTGCACCTGGGCGCT 180
Db 8927 AATTCGAGATCGAATAACAAGTAGAGCAGAGGAGCTTCGAAACACTTGGACCTGGGCGCT 8986
Qy 181 CCTCAGCCAATGATGCCCTGGACTCTCCCTTTTAGGACCTCTAGCAGCTAATAATT 240
Db 8987 CTTCAGCCAATGATGCCCTGGACTCTCCCTTTTAGGACCTCTAGCAGCTAATAATT 9046
Qy 241 TTTTACTCTCTTTGGACCTGTATCTTCAACTCTCTTTGTTAAAGTTCTCTCTCCAGAA 300
Db 9047 GCTACTTCTCTTTGGACCTGTATCTTAACTCTCTTTGTTAACTTTGTTCTCTCCAGAA 9106
Qy 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAAACCCAGATGCAGTCCATGACTAA 360
Db 9107 CGAAGCTGTAAAGCTA-----CAAATGGAGCCCAAGATGCAGTCCCAAGACTAA 9154
Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTGATGCTTAATGACATGA 420
Db 9155 GATCTACACAGACCCCTGGACTGGCTGTAGCCACGATCTGATGTTAATGACATCA 9214
Qy 421 AGTCACCCCTCCGAGGAATCTCACTGCACACCCCTACTACCTCCCAATTCAGTAGG 480
Db 9215 AGGCACCCCTCCTGAGGAATCTCACTGCACACCCCTACTACGCCCCCAATTCAGTAGG 9274
Qy 481 AAGCAGTTAGAGCAGTTGTGTCAGCCAACCTCCCCAACAGTACTTGGGTTTCTCTTGGAGA 540
Db 9275 AAGCAGTTAGAGCAGTTGTGTCAGGCCNACTCCCCAACAGCAGCTTAGGTTTCTCTTGGAGA 9334
Qy 541 GGGTGGACTGAGAGACAGGACTAGCTGGAATTTCTCCTAGGCTGACTAAGAAATCCCAAGCCT 600
```

```
Db 9335 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCTAGGCCTACTAAGAAATCCCTAAGCCT 9394
Qy 601 ANCTGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
Db 9395 AGCTGGAAGGTGACCACTCCATCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 9454
Qy 661 ACCAATCAGAGAGTCTCTAAATGCTAATCAGGCAAAAAACAGGAGGTAAAGCAATAGCC 720
Db 9455 ACCAATCAGAGAGTCTCTAAATGCTAATCAGGCAAAAAACAGGAGGTAAAGCAATAGCC 9514
Qy 721 AATCATCTATTGCTGAGACAGACGGGAGGAGCAGAGGATTGGGATATAAATCAGGCA 780
Db 9515 AATCATCTATTGCTGAGACAGACGGGAGGAGCAGAGGATTAAGGATATAAATCAGGCC 9574
Qy 781 TTCAAGCCAGCAACAGACCCCTTTGGGTGCCCTCCCATTTGATGGGAGCTCTGTTTT 840
Db 9575 TTCGAGCCAGCAACAGACCCCTTTGGGTGCCCTCCCTTTGATGGGAGCTCTGTTTT 9634
Qy 841 CACTCTATTTCACTCTAATTAATCAATGCAATGCACTCTCTGCTCGTGTGTTTTATGG 900
Db 9635 CATGCTATTTCACTCTAATTAATCTTGCAACTGCACTCTCTGCTCCATCTTTGTTACGG 9694
Qy 901 CTCAAGCTGAGCTTTGTTGCGCCATCCACCACTGCTGTTTGGCCAGCTCACAGACCGCT 960
Db 9695 CTTGAGCTGAGCTTTGCTCGCTGTCCACCACTGCTGTTTGGCGCCACCGCAGACCGGCC 9754
Qy 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGCCACTGTGCTCCTGATCCAGCGAGGT 1020
Db 9755 ACTGACTCCCATCCCTCGGATCTCGCAGGGTGTCACCTGTCTCTGATCCAGCGAGGT 9814
Qy 1021 ACCATTGCCACTCCGATCAGGCTAAAGCTTGCCATTTCTGCTGATGGCTTAAGTGC 1080
Db 9815 GCCATTGCGCGTCCCAATCAGGCTAAAGCTTGCCATTTCTGCTGATCACTAAGTGC 9874
Qy 1081 TGGGTTGCTCTAATAGAACTGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db 9875 CAGGTTGCTCTAATGAGCTGAACACTAGTCACTGGGTTCCAGGTTCTCTCTGTCAC 9934
Qy 1141 CCACGGCTCTAATAGAGCTATAACACTCACCGCATGGGCCAAGATTCCATTCCTGGTA 1200
Db 9935 CCACAGCTCTAATAGAGCTATAACACTCACCATGATGGGCCAAGTTCCATTCCTT-GAA 9993
Qy 1201 TCTGTAGGCCAAGAACCCAGGTGAGAGANGTGGGCTTGCCACCATTTGGGAAGTGG 1260
Db 9994 TCCATGAGGCCAAGAACCCAGGTGAGAGAACATGAGGCTTGCCACCATCTTGGAGCAG 10053
Qy 1261 CCACATGCGATTTTGGTAGCGGCCACCACTCTTGGGAGCTGTGGGAGCAAGGATCCC 1320
Db 10054 CCGGCCACCATCTTGGAGCAGCTCACCCATCTTTGGGAGCTCTGTGAGCAAGGACCCC 10113
Qy 1321 CCAGTAAAC 1329
Db 10114 CAAGTAAAC 10122

RESULT 6
AC133134
LOCUS AC133134 148711 bp DNA linear PRI 28-NOV-2002
DEFINITION Homo sapiens chromosome 5 clone RP11-120B7, complete sequence.
ACCESSION AC133134
VERSION AC133134.2 GI:25815299
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 148711)
REFERENCE
  1 DOE Joint Genome Institute and Stanford Human Genome Center.
  Direct Submission
  TITLE
  JOURNAL
  AUTHORS
  UNPUBLISHED
  2 (bases 1 to 148711)
REFERENCE
  2 DOE Joint Genome Institute and Stanford Human Genome Center.
  AUTHORS
```

```
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2002) DOE Joint Genome Institute, 2800 Mitchell
  Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 148711)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (28-NOV-2002) DOE Joint Genome Institute, 2800 Mitchell
  Drive, Walnut Creek, CA 94598, USA
COMMENT On Nov 28, 2002 this sequence version replaced gi:22758291.
  Draft Sequence Produced by DOE Joint Genome Institute
  www.jgi.doe.gov
  Finishing Completed at Stanford Human Genome Center
  www.shgc.stanford.edu
  Quality: Phrap Quality >=40 99.9% of Sequence;
  Estimated Total Number of Errors is 0.7.
FEATURES
  source
    1. 148711
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="5"
      /clone="RP11-120B7"
ORIGIN
Query Match 80.3%; Score 1067; DB 9; Length 148711;
Best Local Similarity 88.9%; Pred. No. 0;
Matches 1175; Conservative 0; Mismatches 143; Indels 4; Gaps 2;
Qy 1 TCAAAATCGAAGAGCTTTAGACTTGCTTAACCGCCAAAGAGGGGAACTGTTTATTTT 60
Db 99272 TCAAAATCGAAGAGCTTTAGACTTGCTTAACCGCCAAAGAGGGGAACTGTTTATTTT 99331
Qy 61 AGGGGAGAGATGCTGTAGTATGTTAATCAATCTGGAATCATTTACTGAGAAAGTTAAAGA 120
Db 99332 AAGGAGAGATGCTGTATTAATGTTAATCAATCCAGATCGTCACTGAGAAAGTTAAAGA 99391
Qy 121 AATTGAGATCGAATATATATGATAGAGAGAGACCTTCAAAACACTGCACCCCTGGGCGCT 180
Db 99392 AATTGAAATCAAAATCAACGTAAGCAAAAGAGCTTCAAAACGTGGGACACTGAGCGCT 99451
Qy 181 CCTCAGCAATGATGATCGCTGGAGCTCTCCCTCTTTAGGACCTCTAGCAGCTATAATTT 240
Db 99452 CCTCAGCAATGATGATCGCTGGAGCTTTCCCTCTTTAGGACCTCTAGCAGCTATAATTT 99511
Qy 241 TTTACTCTCTTTTGGACCCCTGTATCTTCAACTCTTCTTGTAAAGTTGTCTCTCCAGAT 300
Db 99512 GCTACTCAITTTTAGACCCCTGTATCTTAACTCTCTTGTAAAGTTGTCTCTCCAGAT 99571
Qy 301 TGAAGCTGTAAGCTACAAATAGTCTTCAATGGAACCCCAAGATGAGTCCATGACTAA 360
Db 99572 CGAAGCTGTAAGCTACAAATCAATCTTCAAAATGGAGCCCAAGATGAGTCCATGACTAA 99631
Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
Db 99632 GATCTACCGCGGACCCCTGGAGCTGGCTGCTAGCCATGCTCTGATGTTAATGACATTGA 99691
Qy 421 AGTCACCCCTCCCGAGGAAATCTCAACTGCAACCCCTACTACACTCCCAATTCAGTAGG 480
Db 99692 AGGCACCCCTCCAGAGGAAATTTCACTGCAACCCCTACTACACCCCAATTCAGCAGG 99751
Qy 481 AAGCAGTTAGAGAGTTGTCAGCAACCTCCCAACAGTACTTTGGGTTTCTGTTGAGA 540
Db 99752 AAGCAGTTAGAGTTGTCGTCGGCAACCTCCCAACAGCAGTCTAGGTTTCTGTTGAGA 99811
Qy 541 GGGTGAAGTGAAGAGCAGGACTAGCTGGATTTCTAGGCTGACTAGAAATCCCAAGGCT 600
Db 99812 GGGGTTACTGAGAGACAGGACTAGCTGGATTTCTAGGCGGACTAGAAATCCCTAAGCCT 99871
Qy 601 ANCTGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
Db 99872 AGCTGGAAGGTGACCGCATCCACCTTTAAACATGGGGCTTGCAACTTAGCTCACACCTG 99931
Qy 661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAAAACAGGAGGTAAAGCAATAGCC 720
```

[illegible]

```
Qy 481 AAGCAGTTAGAGCAGTGTGTGAGCCAACTCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
Db 9275 AAGCAGTTAGAGCAGTGTGTGAGCCAACTCCCAACAGTACTTGGGTTTTCCTGTTGAGA 9334
Qy 541 GGGTGGACTCAGACAGGAGTACTGAGTATTCCTAGGCTGACTAAGAAATCCCNAGCCT 600
Db 9335 TGGGGAGTCTAGACAGGAGTACTGAGTATTCCTAGGCTGACTAAGAAATCCCNAGCCT 9394
Qy 601 ANCTGGGAAGTGAACCGTACCTCCTTTTAAACATGGGGCTTGGCACTTAGCTCACACCCG 660
Db 9395 AGCTGGGAAGTGAACCGTACCTCCTTTTAAACATGGGGCTTGGCACTTAGCTCACACCTG 9454
Qy 661 ACCAATCAGAGAGTCTCTAAATGCTAATCAGGCAAAAAACAGAGGTAAGCAATAGCC 720
Db 9455 ACCAATCAGAGAGTCTCTAAATGCTAATCAGGCAAAAAACAGAGGTAAGCAATAGCC 9514
Qy 721 AATCATCTATTGCTGAGACAGACGGGAGGAGGATTCGGATATAAATCAGGCA 780
Db 9515 AATCATCTATTGCTGAGACAGACGGGAGGAGGATTCGGATATAAATCAGGCC 9574
Qy 781 TTCAAGCCAGCAACAGCAACCCCTTTTGGGT--CCCTCCCATTTGATGGGAGCTCTGTT 838
Db 9575 TTCGAGCCAGCAACAGCAACCCCTTTTGGGTCCCTCCCTCTTGTATGGGAGCTCTGTT 9634
Qy 839 TTCACTCTATTCTACTTATAATCATGCAACTGCACTCTCTGTCCTGTTTAT 898
Db 9635 TTCATGCTATTCTACTTATAATCATGCAACTGCACTCTCTGTCCTGTTTAT 9694
Qy 899 GGCTCAGCTGAGCTTTGTTGCCATCCACCACTGCTGTTGCCACCGTCACAGACCCG 958
Db 9695 GGCTTGAGCTGAGCTTTGTTGCCATCCACCACTGCTGTTGCCACCGTCACAGACCCG 9754
Qy 959 CTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCTGATCAGCGAG 1018
Db 9755 CCATGACTCCATCCCTCCGGATCCTGAGGGTGTCCACTGTGTTCTGATCCAGCGAG 9814
Qy 1019 GTACCAATGCCACTCCGATCAGGTAAGAGCTTGCATGTTCTGTCGATAGGCTAAGTG 1078
Db 9815 GTGCCAATGCCCTCCCAATCAGGCTAAGAGCTTGCATGTTCTGTCGATCATAAGTG 9874
Qy 1079 CTTGGTCTTCTCTATAGACTGACACTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1138
Db 9875 TCAGGTTCTGCTCTCTATAGACTGACACTGTCCTGCTGCTGCTGCTGCTGCTGCTG 9934
Qy 1139 ACCACGGCTTCTAATAGAGCTATAACACTCAGCGATGCCCAAGATTCATTCCTTGG 1198
Db 9935 ACCACAGCTTCTAATAGAGCTATAACACTCAGCATGCCCAAGTTCATTCCTT-G 9993
Qy 1199 TATCTGTAGGCAAGAACCCCGTTCAGAGANGTGGCTTGCACCAATTTGGGAAGT 1258
Db 9994 AATCCATGAGGCAAGAACCCCGTTCAGAGANGTGGCTTGCACCAATTTGGGAAGC 10053
Qy 1259 GGCCCACTGCAATTTGGTAGCGGCCACCAACCATCTTGGAGCTGTTGGAGCAAGATC 1318
Db 10054 AGCCCGCCACCATCTTGGAGAGAGCTCACCACCATCTTGGAGCTGTTGGAGCAAGATC 10113
Qy 1319 CCCAGTAACA 1329
Db 10114 CCCAGTAACA 10124
```

```
RESULT 8
AF520477S2 2694 bp DNA linear PRI 11-FEB-2004
LOCUS Homo sapiens individual 132 allele A, envelope glycoprotein gene,
DEFINITION complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION AF520478
VERSION AF520478.1 GI:33410924
KEYWORDS 2 of 2
SEGMENT Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 2694)
Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
Bonnaud,B., Lucotte,G., Duret,L., and Mandrand,B.
The endogenous retroviral locus ERVW1 is a bona fide gene involved
in hominoid placental physiology
Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
14757826
REFERENCE 2 (bases 1 to 2694)
Mallet,F., Bouton,O. and Oriol,G.
Direct Submission
Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
CNRS-BioVerieux, Ecole Normale Supérieure de Lyon - 46 allée
d'Italie, Lyon 69364 cedex 07, France
Location/Qualifiers
1. .2694
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="individual 132 allele A"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-q22"
/sex="female"
/cell_type="PBMC's"
/notes="Caucasian
endogenous_virus: HERV-W"
misc_feature 43..57
/notes="splice acceptor site"
CDS 284..1900
/notes="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAQ17561.1"
/db_xref="GI:33410925"
/translation="MALPYHILFLFTVLLPSFTLTAPPCCMTSSPYQSFRLMRMORP
GNIDAFYSLSKSGTFTFTAHMPNCHVSATLKNHNTYTGMINSPSCGGGLV
TCTWYFTQMGSDGGVQDQAREKHVEISQITRVHGTSSPYKGLDLKSHETLRT
HRLVSLFNLTITLHEVSAQNTNCWICLPLNFRPVSIPIPEQWNFETHTTSV
LVGLPVNLLEITHTSNLTCVKFSNTTNTNSQCIWPTQIVCLPSGFFVCGTSA
YCLNGSSSMCFSLSPQMTIYTBQDLYNYVISKPRKRVPILEFVIGAVLGALG
TGIGITTTSTQFVYKLSQMDGMERVADSLVTLQDLSLAAVLQNRALDLTAE
RGCTCLFLBEECCYYNQSGIVTEKVEIRDRLOREBELRNTGPGWLLSOWMPILP
FGLPLAAILLLFGLFCIFNLLVNFVSSRIEAVLQWEPKQSKTKIYRPLDRPAS
RSDVNDIKGTPPPEISAAQLRPNAGSS"
misc_feature 1705..1720
/notes="splice acceptor site"
LTR 1959..2694
Query Match 78.0%; Score 1036.8; DB 9; Length 2694;
Best Local Similarity 90.2%; Pred. No. 1e-304;
Matches 1134; Conservative 0; Mismatches 110; Indels 13; Gaps 2;
Qy 1 TCAAAATCGAGAGCTTAGACTTGTCTAACCGCAAGAGGGGGACCTGTTATTTT 60
Db 1423 TCAAAATCGAGAGCTTAGACTTGTCTAACCGCAAGAGGGGGACCTGTTATTTT 1482
Qy 61 AGGGGAAGATGCTGTAGTATCTTAATCAATCTGGAATCATTTACTGAGAAAGTTAAAGA 120
Db 1483 AGGGGAAGATGCTGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1542
Qy 121 AATTTGAGATCGAATAATGTTAGACAGAGGACCTTCAAAAACATGACCTGGGCGCT 180
Db 1543 AATTTGAGATCGAATAACACGTAGACAGAGGAGCTTGAAGACATGACCTGGGCGCT 1602
Qy 181 CCTCAGCAATGATGCCCTGGACTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATT 240
Db 1603 CCTCAGCAATGATGCCCTGGACTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATT 1662
Qy 241 TTTACTCCTTTTGGACCTGTATCTTCAACTTCTTGTAAAGTTGTCTCTCCAGAAT 300
Db 1663 GCTACTCCTTTTGGACCTGTATCTTCAACTTCTTGTAAAGTTGTCTCTCCAGAAT 1722
```

```

QY 301 TGAAGCTGTAAGCTACAAATAGTTCTTCAATGGACCCAGATCGAGTCCATGACTAA 360
Db 1723 CGAAGCTGTAAACTA-----CAAATGGAGCCCAAGATGAGTCCCAAGACTAA 1770
QY 361 AATCTACCGTGGACCCCTGACCGGCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
Db 1771 GATCTACCGCAGACCCCTGACCGGCTGCTAGCCAGATCTGATGTTAATGACATCA 1830
QY 421 AGTCACCCCTCCGAGGAAATCTCAACTGACAAACCCCTACTACACTCTCAATTCAGTAGG 480
Db 1831 AGGCACCCCTCTCTAGGAAATCTCAGCTGACAAACCTCTACTACGCCCAATTCAGCAGG 1890
QY 481 AAGCAGTTAGCAGAGTTGTGACGCCAACCCTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 540
Db 1891 AAGCAGTTAGAGCGGTGCTCGGCCAACCTCCCAACAGCACTTAGGTTTCTCTGTTGAGA 1950
QY 541 GGGTGAAGTACAGAGACAGGACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCNAGCCCT 600
Db 1951 TGGGGAGTACAGAGACAGGACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCNAGCCCT 2010
QY 601 ANCTGGGAAGTGAACCCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
Db 2011 AGCTGGGAAGTGAACCATCCACCTTTAAACAGCGGGCTTGCAACTTAGCTCACACCTG 2070
QY 661 ACCAATCAGAGAGCTCACTAAATGCTTAATCAGGCAAAACAGGAGTAAAGCAATAGCC 720
Db 2071 ACCAATCAGAGAGCTCACTAAATGCTTAATTAGGCAAAACAGGAGTAAAGCAATAGCC 2130
QY 721 AATCATCTATTGCTGAGACACAGCGGGAAGACAGGAGTGGGATATAAATCTCAGGCA 780
Db 2131 AATCATCTATTGCTGAGACACAGCGGGAAGACAGGAGTGGGATATAAATCTCAGGCA 2190
QY 781 TTCAAGCAGCAACAGCAACCCCTTTGGGTGCCCTCCCATTTGATGGGAGCTCTGTTTT 840
Db 2191 TTCAAGCAGCAACAGCAACCCCTTTGGGTGCCCTCCCATTTGATGGGAGCTCTGTTTT 2250
QY 841 CATCTATTTCACCTATTAAATCATGCAACTGACCTCTCTGCTCGGTGCTTTTATGG 900
Db 2251 CATGCTATTTCACCTATTAAATCTTGAACATGCACTCTCTGCTCGGTGCTTTTATGG 2310
QY 901 CTCAGCTGAGCTTTGCTCGGCATCCACCACTGCTGTTGGCCACCGTCAACAGCCGCT 960
Db 2311 CTGAGCTGAGCTTTGCTCGGCATCCACCACTGCTGTTGGCCACCGTCAACAGCCGCT 2370
QY 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGCTCACTGCTGCTCTGATCCAGCGAGGT 1020
Db 2371 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGCTCACTGCTGCTCTGATCCAGCGAGG 2430
QY 1021 ACCATTGCCATCCCGATCAGGCTAAAGGCTTGCCATTTGCTGATGCTAAAGTGC 1080
Db 2431 ACCATTGCCATCCCGATCAGGCTAAAGGCTTGCCATTTGCTGATGCTAAAGTGC 2490
QY 1081 TGGGTTGCTTAATAGAACTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db 2491 TGGGTTGCTTAATAGAACTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2550
QY 1141 CCACGGCTTCTAATAGAGCTATTAACACTCAGCGCATGCGCCCAAGATTCCATTCCTTGGTA 1200
Db 2551 CCACGGCTTCTAATAGAGCTATTAACACTCAGCGCATGCGCCCAAGATTCCATTCCTTGGTA 2609
QY 1201 TCTGTGAGGCAAGAACCCCGAGTACAGAGANGTGGGCTTGCCACCATTTGGGAAG 1257
Db 2610 TCCATAAGGCCAAGAACCCCGAGTACAGAGANGTGGGCTTGCCACCATTTGGGAAG 2666

```

RESULT 9

AF52048982

LOCUS

DEFINITION Homo sapiens individual 22 allele A, envelope glycoprotein gene, complete cds, and 3' long terminal repeat, complete sequence.

ACCESSION AF520490

VERSION AF520490.1

KEYWORDS AF520490.1 GI:33410948

```

SEGMENT 2 of 2
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2694)
AUTHORS Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G., Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
TITLE The endogenous retroviral locus ERVW1 is a Don't fide gene involved in hominoid placental physiology
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
PUBMED 14757826
REFERENCE 2 (bases 1 to 2694)
AUTHORS Mallet,F., Bouton,O. and Oriol,G.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2002) Retrovirology Department, UMR 2142 CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée d'Italie, Lyon 69364 cedex 07, France
FEATURES
    source
        1..2694
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /isolate="individual 22 allele A"
            /db_xref="taxon:9606"
            /chromosome="7"
            /map="7q21-q22"
            /sex="male"
            /cell_type="PBMC's"
            /notes="Caucasian"
            endogenous_virus: HERV-W
        misc_feature
            43..57
                /notes="splice acceptor site"
        CDS
            284..1900
                /notes="syncytin"
                /codon_start=1
                /product="envelope glycoprotein"
                /protein_id="AAQ17567.1"
                /db_xref="GI:33410949"
                /translation="MALPVIHFLFTVLLPSFTLTAPPKRCMTSSSYQBFWMQRP
                GNIDAPYSRLSGTPTTAHTHPRNCVHSATLCMHANTHYWTGMNPSCPGGLV
                TFCWTYFTGTGSDGGVDOAREKHVKEVISLTTRVHGTSSPYKGLDLSKLEHETLR
                HTPLVSLFWTLTLGLHEVSAONPTNCWICLPLNFRPVSIPTVPEOWNNFEINTTSTV
                LVGPLVSNLEITHSNLTCVKPNTTYTNSQCIRWVTPTQIVCLPSFFVCGTSA
                YRLNGSSSMSCFLPVPNTTYTQDLYNVIYISKPRKRPFLPFIIVGAGVGLAG
                TGCTGTTTQFYKLSQELNGMDVADSLVTLQDLNLAALVQLNRALDILLTAE
                RGCTCLFLGEECCYVNGSVIETKVEIKRDIQRRAELRNTGPMGLLSQMMFWILP
                FLGPLAAIILILLFGPCIFNLVNFVSSRIEAVKLQWEPKMQSKTKIYRRPLDRPAS
                RSDVNDIKGTPPEISNAQPLLRPSAGSS"
        misc_feature
            1705..1720
                /notes="splice acceptor site"
            1959..2694

```

Query Match 78.0%; Score 1036.8; DB 9; Length 2694;
 Best Local Similarity 90.2%; Pred. No. 1e-304;
 Matches 1134; Conservative 0; Mismatches 110; Indels 13; Gaps 2;

```

QY 1 TCAAAATCGAAGAGCTTTAGACTTGTCAACCGCCAAAGAGGGGGAACCTGTTTATTTT 60
Db 1423 TCAAAATCGAAGAGCTTTAGACTTGTCAACCGCTGAAGAGGGGGAACCTGTTTATTTT 1482
QY 61 AGGGGAAGAAATGCTGTTAGTATGTTAATCAATCTGGAATCATTAATGAGAAAGTAAAGA 120
Db 1483 AGGGGAAGAAATGCTGTTATGTTAATCAATCGGAATCGTCACTGAGAAAGTAAAGA 1542
QY 121 AATTGAGATCGAATAATAATGATAGCAGAGGACCTTCAAAACACTGCACCTGGGCGCT 180
Db 1543 AATTGAGATCGAATAACAAGTAGCAGAGGAGCTTCGAAACACTGCACCTGGGCGCT 1602
QY 181 CCTCAGCAATGATGCTCCCTGGAGCTCTCCCTCTTAGGACCTCTAGCAGCTATAATT 240
Db 1603 CCTCAGCAATGATGCTCCCTGGAGCTCTCCCTCTTAGGACCTCTAGCAGCTATAATT 1662

```



```
Db 1603 CCTCAGCCAAATGATGCCCTGGATTCTCCCTCTTTAGGACCTCTTAGCAGCTATATAATTT 1662
Qy 241 TTTACTCCTCTTTGGACCCCTGTATCTTCAACTTCCTTGTAAAGTTGTCTCTTCACGAAT 300
Db 1663 GCTACTCCTCTTTGGACCCCTGTATCTTAACTCTCTTGTAACTTGTCTCTTCACGAAT 1722
Qy 301 TGAAGCTGTAAGCTCAAAATAGTCTTCTTCAATAGGAACCCAGATCGAGTCATGACTAA 360
Db 1723 CGAAGCTGTAAGCTCAAAATAGTCTTCTTCAATAGGAACCCAGATCGAGTCATGACTAA 1770
Qy 361 AATCTACCTGGACCCCTGGACCGGCTGTAGACTATGCTCTGATGTTGTTAAAGACATGA 420
Db 1771 GATCTACCGAGACCCCTGGACCGGCTGTAGCCACGATCTGATGTTAAAGACATGA 1830
Qy 421 AGTCACCCCTCCGAGGAATCTCAACTGCAACACCCCTACTACACTCAATTCAGTAGG 480
Db 1831 AGGCACCCCTCCGAGGAATCTCAGCTGCACAACTCTACTAGCCCCCAATTCAGCAGG 1890
Qy 481 AGCAGTTAGAGAGTGTGAGCAACCTCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
Db 1891 AAGCAGTTAGAGCGGTGCTGGCAACCTCCCAACAGCAGCTTAGGTTTTCCTGTTGAGA 1950
Qy 541 GGGTGACTGAGAGACGAGTACTGGAATTCCTAGGCTGACTAAGAAATCCCAAGCCT 600
Db 1951 TGGGGGACTGAGAGACGAGTACTGGAATTCCTAGGCTGACTAAGAAATCCCAAGCCT 2010
Qy 601 ANCTGGGAAGGTGACCGCATCTCACTTTTAAACATGGGGCTTGCAACTTAGCTCACACCGG 660
Db 2011 AGCTGGGAAGGTGACCACTCACTTTTAAACAGCGGGCTTGCAACTTAGCTCACACCTG 2070
Qy 661 ACCAATCAGAGACTCACTAAATGCTAATCAGGCAAAAACAGGAGTAAAGCAATAGCC 720
Db 2071 ACCAATCAGAGACTCACTAAATGCTAATTAGGCAAAAACAGGAGTAAAGCAATAGCC 2130
Qy 721 AATCATCTATTGCTGAGAGCAGCGGAGGACAGGATTTGGGATATAAAGTCAAGCA 780
Db 2131 AATCATCTATTGCTGAGAGCAGCGGAGGACAGGATATAAAGTCAAGCA 2190
Qy 781 TTCAAGCAGCAACAGCAACCCCTTTGGTCCCTCCCACTTGTATGGGAGCTCTGTTT 840
Db 2191 TTGAGCGCGCAACGCAACCCCTTTGGTCCCTCCCTTTGTATGGGAGCTCTGTTT 2250
Qy 841 CACTCTATTCTACTCTATTAATCATGCAACTGCACTCTTCTGGTCCGTTTATGG 900
Db 2251 CATGCTATTCTACTCTATTAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTACGG 2310
Qy 901 CTCAGCTGAGCTTTGTTGGCCATCCACCACTGCTTTTGCCACCGTCACAGACCGCT 960
Db 2311 CTTGAGCTGAGCTTTGCTGCGCTCCCATCCACCACTGCTTTTGCCGCGCCACCGCAGACCGCC 2370
Qy 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGTATCCAGCGAGGT 1020
Db 2371 GCTGACTTCCATCCCTCTGATCATGAGGTTGCTGCTGCTGCTGCTGATCCAGCGAGGC 2430
Qy 1021 ACCATTGGCACTCCCGATCAGGCTTAAAGCTTGGCAATGTTTCTCGATGGCTAAGTGC 1080
Db 2431 ACCATTGGCGCTCCCAATCGGCTTAAAGCTTGGCAATGTTTCTCGATGGCTAAGTGC 2490
Qy 1081 TGGGTTTGTCTTAATAGAACTGAACACTGTGTGCTGCTGGTTTCCATGTTCTCTTCCATGAC 1140
Db 2491 TGGGTTTGTCTTAATAGAACTGAACACTGTGTGCTGCTGGTTTCCATGTTCTCTTCTGTGAC 2550
Qy 1141 CCAGCGCTCTTAATAGAGCTATAAAGCTCAACCGCATGGCCCAAGATTCCATCTCTTGGTA 1200
Db 2551 CCAGAGCTCTTAATAGAGCTATAAAGCTCAACCGCATGGCCCAAGATTCCATCTCTTGGTA 2609
Qy 1201 TCTGTGAGGCGCAAGAACCCAGGCTCAGAGAAAGTGAGGCTTGCCACCATTTGGGAAG 1257
Db 2610 TCCATAAGGCGCAAGAACCCAGGCTCAGAGAAAGTGAGGCTTGCCACCATTTGGGAAG 2666
```

RESULT 11

```
AF520529S2 2694 bp DNA linear PRI 11-FEB-2004
LOCUS Homo sapiens individual 78 allele A, envelope glycoprotein gene,
DEFINITION complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION AF520530
VERSION AF520530.1 GI:33411028
SEGMENT 2 of 2
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2694)
AUTHORS Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
TITLE The endogenous retroviral locus ERVWE1 is a bona fide gene involved
JOURNAL in hominoid placental physiology
PUBMED Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
14757826
REFERENCE 2 (bases 1 to 2694)
AUTHORS Mallet,F., Bouton,O. and Oriol,G.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée
d'Italie, Lyon 69364 cedex 07, France
FEATURES
source
1..2694
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="individual 78 allele A"
/db_xref="taxon:9606"
/chromosomes="7"
/map="7q21-q22"
/sex="female"
/cell_type="PBMC's"
/note="Caucasian
endogenous_virus: HERV-W"
misc_feature 43..57
/note="splice acceptor site"
284..1900
/note="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAQ17587.1"
/db_xref="GI:33411029"
/translation="MALPYHIFLFTLLPFTLTAPPCCMTSSPYQEFLLWRMQRP
GNIDAPSVRLSKGTPPTFANTHPRNCVHSATLCMHANTHYWTGMINPSCPGGLGV
TVCWYFTOTGSDGGVQDOAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRT
HRLVSLFNTLTGLHEVSAQNPNCWICLPNFRPYVISIPVPEQWNFEINNTSV
LVGPLVSNLEIHTNLCTCFKFSNTTYTTSOCIRWVTPTQIVCLPSGIFVCGTSA
YRLNGSSSEMCFLSFLVPPMTIYTEQDLYNYVISKPRNKRVPILPFVIGAGVIGALG
TGIGGTTSTQFYKLSQELNGDMERVADSLVTLDQLNSLAAVLQNRRLDLLTAE
RGCTCLFLGEECCYVYNQSGIVTEKVKERDRIRORAEELRNTGPKGLLSOMMPWILP
FLGPLAAIITLLLLFGPCIENLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRLDRPAS
RSDVNDIKGTPPEEISAAQPLLRPSAGSS"
misc_feature 1705..1720
/note="splice acceptor site"
1959..2694
LTR
ORIGIN
Query Match 78.0%; Score 1036.8; DB 9; Length 2694;
Best Local Similarity 90.2%; Pred. No. 1e-304;
Matches 1134; Conservative 0; Mismatches 110; Indels 13; Gaps 2;
Qy 1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCAAAAGAGGGGAACTGTTATTTT 60
Db 1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAGAGGGGAACTGTTATTTT 1482
Qy 61 AGGGGAAGAAATGCTGTAGTAGTATGTTAATCAATCTGGAATCATTTACTGAGAAAGTTAAGA 120
Db 1483 AGGGGAAGAAATGCTGTATGTTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAGA 1542
Qy 121 AATTGAGATCGAATATAATGTTAGACAGAGGAGCTTCAAAACACTGCACCTCGGGCCT 180
```

Db 1543 AATTCGAGATCGAATAACAACGATAGACGAGGAGCTTCGAAACACTGGACCTGGGCGCT 1602
Qy 181 CCTCAGCAATGGATGCCCTGGACTCTCCCTCTTAGGACCTCTAGCAGCTATAATATT 240
Db 1603 CCTCAGCAATGGATGCCCTGGACTCTCCCTCTTAGGACCTCTAGCAGCTATAATATT 1662
Qy 241 TTACTCTCTTTGGACCCCTGTATCTTCAACTCTCTTTGTTAAAGTTTCTCTTCCAGAA 300
Db 1663 GCTACTCTCTTTGGACCCCTGTATCTTAACTCTCTTTGTTAACTTTCTCTTCCAGAA 1722
Qy 301 TGAAGCTGTAAGCTACAATAGTTCTTCAATGGAAACCCAGATGCAATGCAATGACTAA 360
Db 1723 CGAAGCTGTAAACTA-----CAAATGGAGCCCAAGATGCAATGCAAGACTAA 1770
Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTGTGATGCTTAATGACATTGA 420
Db 1771 GATCTACCGCAGACCCCTGGACCGGCTGTAGCCACGATCTGATGTTAATGACATCAA 1830
Qy 421 AGTCACCCCTCCGAGGAAATCTCAACTGCACAAACCCCTACTACACTCCAATTCAGTAGG 480
Db 1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACTCTACTACGCCCAATTCAGCAGG 1890
Qy 481 AAGCAGTTAGACGAGTTGTGAGCAACCTCCCAACGATCTTGGGTTTCTCTGTTGAGA 540
Db 1891 AAGCAGTTAGAGCGGTGCTCGGCCAACCTCCCAACAGCAGCTTAGGTTTCTCTGTTGAGA 1950
Qy 541 GGGTGGACTCAGACAGGACTAGCTGATTTCTAGGCTGACTAAGAATCCCNAGACCT 600
Db 1951 TGGGGACTCAGACAGGACTAGCTGATTTCTAGGCTGACTAAGAATCCCTAGACCT 2010
Qy 601 ANCTGGAAGGTGACCGACTCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
Db 2011 AGCTGGAAGGTGACCAATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070
Qy 661 ACCAATCAGAGCTCACTAAATGCTAATCAGGCAAAAAACAGAGGTAAGCAATAGCC 720
Db 2071 ACCAATCAGAGCTCACTAAATGCTAATAGGCAAAAAACAGAGGTAAGCAATAGCC 2130
Qy 721 AATCATCTATTGCTGAGACACAGCGGGAAGGACAGGATTCGGATATAAATCAGGCA 780
Db 2131 AATCATCTATTGCTGAGACACAGCGAGGAGCAATGATCGGATATAAACCAGTGC 2190
Qy 781 TTCAAGCAGCAACAGCAACCCCTTTGGGTCCCTCCCACTGTATGAGGAGCTCTGTTT 840
Db 2191 TCGAGCGGCAACGGCAACCCCTTTGGGTCCCTCCCTTTGTATGGGAGCTCTGTTT 2250
Qy 841 CACTCTATTCTACTTATTAATCATGCACTGCACTCTTCTGGTCGGTGTGTTTATGG 900
Db 2251 CATGCTATTTCACTTATTAATCTTGCAACTGCACTCTTCTGGTCCATGTTCTTACGG 2310
Qy 901 CTCAAGCTGAGCTTTGTTTCGCATCCACCACTGCTGTTTGGCCACGTCACAGACCGCT 960
Db 2311 CTGAGCTGAGCTTTGCTCGCAATCCACCACTGCTGTTTGGCCACGTCACAGACCGCT 2370
Qy 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGCTCACTGTGCTGCTGATCCAGCAGGT 1020
Db 2371 GCTGACTTCCATCCCTCTGATCATGCAAGGTGTGCGCTGTGCTCTGATCCAGCAGGC 2430
Qy 1021 ACCATGTCACCTCCCGATCAGGCTAAGCGTTGCCATTTGTTCTGCTGATGCTAAGTGC 1080
Db 2431 ACCATGTCGCGCTCCCAATCGGCTAAGGCTTGCCATTTGTTCTGCTGATGCTAAGTGC 2490
Qy 1081 TGGGTTGTTCTAATAGAACTGTAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db 2491 TGGGTTCACTCTAATGAGCTGAACACTAGTCACTGGTTCCATGGTTCTCTCTGCTGAC 2550
Qy 1141 CCACGGCTTCTAATAGAGCTATAACACTCAACGATGCGCCCAAGATTCATTCCTTGGTA 1200
Db 2551 CCACAGCTTCTAATAGAGCTATAACACTCAACGATGCGCCCAAGGTTCCATTCCTT-GAA 2609
Qy 1201 TCTGTAGGCGCAAGACCCAGCTCAGAGAGTGGGCTTGCCACCATTTGGGAAG 1257

Db 2610 TCCATAAGGCCAAGAACCCAGCTCAGAGAACAGGAGCTTGCCACCATCTTGGGAG 2666
RESULT 12
LOCUS AF520483S2 2694 bp DNA linear PRI 11-FEB-2004
DEFINITION Homo sapiens individual 147 allele B, envelope glycoprotein gene, complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION AF520484
VERSION AF520484.1 GI:33410936
KEYWORDS 2 of 2
SEGMENT Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 2694)
AUTHORS Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G., Bonnaud, B., Lucotte, G., Duret, L. and Mandrand, B.
TITLE The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
PUBMED 14757826
REFERENCE 2 (bases 1 to 2694)
AUTHORS Mallet, F., Bouton, O. and Oriol, G.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2002) Retrovirology Department, UMR 2142 CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée d'Italie, Lyon 69364 cedex 07, France
FEATURES
source
1..2694
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="individual 147 allele B"
/db_xref="taxon:9606"
/chromosomes="7"
/map="7q21-q22"
/sex="female"
/cell_type="PBMC's"
/note="endogenous_virus: HERV-W"
misc_feature 43..57
/notes="splice acceptor site"
CDS 284..1900
/notes="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAQ17564.1"
/db_xref="GI:33410937"
/translation="MALPYHIFLFTVLLPSFLTAPPCCRCMTSSPYOEFLLWRMRP
GNIDAPYSRLSGKGTPTTAHTMPRNCYHSATLCMHANTHYWTGRMNPSPCGGLV
TVCVTYTOTGMSDGGQVODQAREKHVEVIGSOLTRVHGTSSPYKGLDLSKLHETLR
HVLVSLPNTLTGLHVSANPTNCWICLPLNFRPVSIPIVPEOWNFSTEINTTSSA
LVGPLVNLNLTHTSNLTCVKSNTTYTNSQICIRWVTPPTQIVCLPSGIFVCGTSA
YKLGSSSSMCLSLFLVPPMTIYTEODLYSYVISKPRNKFVILPFVIGAGVGLGALG
TGIGGTTSTQFYKLSQELNGDMERVAADSLVTLQDLNSLAVALVONRRALLLAE
RGCTCLFGECCYVNGQIVTEKVEIRDRIOREAEELRNTGPMGLSSQMPWTLIP
FLGPLAAIILLLLFGPCIFNLVNVFSSRIEAVKLQMEPKMQSKTIYRPLDRPASP
RSDVNDIKGTTPPEISAAQPLLRPNSSGSS"
misc_feature 1705..1720
/note="splice acceptor site"
LTR 1959..2694
ORIGIN
Query Match 77.9%; Score 1035.2; DB 9; Length 2694;
Best Local Similarity 90.1%; Pred. No. 3.1e-304;
Matches 1133; Conservative 0; Mismatches 111; Indels 13; Gaps 2;
Qy 1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCAAAAGAGGGGAACTGTTATTTT 60
Db 1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAGAGGGGAACTGTTATTTT 1482
Qy 61 AGGGGAAGATGCTGTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAGA 120

```
Db 1483 AGGGAAGATGCTGTTATTATGTTAATCAATCGGAATCGTCACTGAGAAAGCTTAAGA 1542
Qy 121 AATTTGAGATCGAATATAATGTAGAGAGAGAGACTTCAAAAACATCGCAACCTGGGGCCT 180
Db 1543 AATTCGAGATCGAATACAACGTAGAGAGAGAGAGCTTCGAAACACATGGACCTGGGGCCT 1602
Qy 181 CCTCAGCAATGATGAGCCCTGGAGCTCTCCCTCTTTAGGACCTCTAGCAGCTATAATATT 240
Db 1603 CCTCAGCAATGATGAGCCCTGGAGTTCCTCCCTCTTTAGGACCTCTAGCAGCTATAATATT 1662
Qy 241 TTTACTCTCTTTTGGACCTGTATCTTCAACTTCTTGTAAAGTTGTCTCTTCCAGAAAT 300
Db 1663 GCTACTCTCTTTTGGACCTGTATCTTCAACTTCTTGTAAAGTTGTCTCTTCCAGAAAT 1722
Qy 301 TGAAGCTGTAAGCTACAAATAGTCTTCAAAATGGAACCCAGAGATGAGTCCATGACTAA 360
Db 1723 CGAAGCTGTAAGCTA-----CAAATGGAGCCCAAGATGAGTCCCAAGACTAA 1770
Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGATGTTAATGACATTTGA 420
Db 1771 GATCTACCGCAGACCCCTGGACCGGCTGTAGCCACGATCTGATGTTAATGACATCAA 1830
Qy 421 AGTCACCCCTCCGAGGAATCTCAACTGCACAAACCCCTACTACACTCCCAATTCAGTAGG 480
Db 1831 AGGCACCCCTCCTGAGGAATCTCAGCTGCACAACTCTACTAGCCCCCAATTCAGCAGG 1890
Qy 481 AAGCAGTTAGAGAGATGTTGAGCAACCTCCCAACAGTACTTGGGTTTCTGTTGAGA 540
Db 1891 AAGCAGTTAGAGCGGTCTGTCGGCCAACTCCCAACAGCAGCTTAGGTTTCTGTTGAGA 1950
Qy 541 GGTGGACTGAGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAGAAATCCCNAGCCT 600
Db 1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAGAAATCCCNAGCCT 2010
Qy 601 ANCTGGAAGGTGACCGCATCTCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
Db 2011 AGCTGGGAAGGTGACCACTCCACCTTTAAACAGCGGGCTTGCAACTTAGCTCACACCTG 2070
Qy 661 ACCAATCAGAGAGCTCACTAAATATGCTAATCAGGCAAAAAACAGAGGTAAAGCAATAGCC 720
Db 2071 ACCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCC 2130
Qy 721 AATCATCTATTGCTGAGACACAGCGGGAAGGACAGGATTTGGGATATAAATCTCAGGCA 780
Db 2131 AATCATCTATTGCTGAGACACAGGAGGAGGACAAATGATCGGGATATAAATCCCAAGTC 2190
Qy 781 TTCAAGCCAGCAACAGCAACCCCTTTGGGTGCCCTCCCAATGTATGGGAGCTCTGTTTT 840
Db 2191 TTGAGCCGCAACAGGCAACCCCTTTGGGTGCCCTCCCTCTTGTATGGGAGCTCTGTTTT 2250
Qy 841 CACTCTATTTCACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTTGTTTTATGG 900
Db 2251 CATGCTATTTCACTCTATTAAATCTTTGCAACTGCACTCTTCTGGTCCATGTTCTTACGG 2310
Qy 901 CTCAAGCTGAGCTTTGTTGGCCATCACCACCTGCTGTTTGGCACCGTCAACAGCCGCT 960
Db 2311 CTTGAGCTGAGCTTTGCTCGCCATCCACCACTGCTGTTTGGCGCCACCCGACAGCCGCGC 2370
Qy 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCACCTGTGCTCCTGATCCAGCGAGGT 1020
Db 2371 GCTGACTTCCATCCCTCTGGATCATGCAAGGTGTCCGCTGTGCTCTGATCCAGCGAGGC 2430
Qy 1021 ACCATTGGCCACTCCCGATCAGGCTAAAGGCTTGCCATTTGTTCTGTCATGGCTAAAGTGC 1080
Db 2431 ACCATTGGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTTGTTCTGTCATGGCTAAAGTGC 2490
Qy 1081 TGGGTTTGTCTTAATAGAACTGAACACTGGTCTCATGGGTTCCATGGTTCTCTTCCATGAC 1140
Db 2491 TGGGTTTCACTTAATGAGCTGAACACTAGTCACTGGGTTCCATGGTTCTCTTCTGTGAC 2550
Qy 1141 CCAAGGCTTCTAATAGAGCTTAACACTCAAGCATGGCCCAAGATTCCAATTCCTTGGTA 1200
Db 2551 CCACAGCTTCTAATAGAGCTTAACACTCAAGCATGGCCCAAGGTTCCATTCCTT-GAA 2609
```

```
Qy 1201 TCTGTGAGCCCAAGACCCAGCTCAGAGAAAGTGAGGCTTGCCACCATTTGGGAAG 1257
Db 2610 TCCATAAGGCCCAAGAACCCAGCTCAGAGAAACGAGGCTTGCCACCATTTGGGAG 2666

RESULT 13
AF520487S2 2694 bp DNA linear PRI 11-FEB-2004
LOCUS Homo sapiens individual 148 allele B, envelope glycoprotein gene,
DEFINITION complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION AF520488
VERSION AF520488.1 GI:33410944
KEYWORDS
SEGMENT
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2694)
AUTHORS Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
TITLE The endogenous retroviral locus ERVWE1 is a bona fide gene involved
JOURNAL in hominoid placental physiology
PUBMED Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
14757826
REFERENCE 2 (bases 1 to 2694)
AUTHORS Mallet,F., Bouton,O. and Oriol,G.
JOURNAL Direct Submission
TITLE Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée
d'Italie, Lyon 69364 cedex 07, France
FEATURES
source
1..2694
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="individual 148 allele B"
/db_xref="taxon:9606"
/map="7q21-q22"
/sex="male"
/cell_type="PBMC's"
/note="Caucasian"
endogenous_virus: HERV-W
misc_feature 43..57
/note="splice acceptor site"
CDS 284..1900
/note="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAQ17566.1"
/db_xref="GI:33410945"
/translation="MALPYHIFLFTVLLPSFTLTAPPCCMTSSSPYQBFLLWRMORP
GNIDAPSYLSKGTPTTAHTHMPNCVHSATLCMHANTHYWTGMINPSCPGGLGV
TVKWTFTQGTGSDGGVQDQAREKHVEISQLTRVHGTSYSPYKGLDLSKLHETLRT
HRLVSLFNTLTGLHEVSAQNTNCWICLPLNRPYVSPVPEQWNFTSINTSV
LVGPLVNSLEITHSNLTCKFNTTNTTNSQICRWPTPTQVLPSPGFVCGTSA
YRLGNSSESMCFSLFPLVPMITTYTDQDLYNNVSKPRKRVPLIPFVIGALG
TGIGGTTTQFYKLSQELNGDMRVADSLVTLOQLNSLAALVQLNRALDLTAE
RGCTCLFLAGECCVYVNGSIIVTEKVEIKDRIORRAEELRNTGPGLLSQMWPILP
FLGPLAAIILLFLFGPCIFNLNVFVSSRIEAVLQNEPKQSKTKIYRPLDRPASP
RSDVNDIKGTPPPEISAAQPLLRPNAGSS"
misc_feature 1705..1720
/note="splice acceptor site"
LTR 1959..2694
ORIGIN

Query Match 77.9%; Score 1035.2; DB 9; Length 2694;
Best Local Similarity 90.1%; Pred. No. 3.1e-304;
Matches 1133; Conservative 0; Mismatches 111; Indels 13; Gaps 2;
Qy 1 TCAAAATCGAGAGCTTTAGACTTGTAAACCGCAAGAGGGGACCTGTTATTTT 60
```

Db 1423 TCAAAATCGAAGAGCTTTAGACTTGTCTAAACCGCTGAAGAGGGGAAACCTGTTATTTT 1482
Qy 61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTTACTGAGAAGTTAAAGA 120
Db 1483 AGGGGAAGAATGCTGTTATTTATGTTAATCAATCGGAAATCGTCACTGAGAAGTTAAAGA 1542
Qy 121 AATTGAGATCGAATAAATGTTAGAGCAGAGGACCTTTCAAAACACATGCAACCTGGGCGCT 180
Db 1543 AATTGAGATCGAATAAATGTTAGAGCAGAGGAGCTTCGAAACACACTGGACCTGGGCGCT 1602
Qy 181 CCTCAGCAATGATGATGCGCTGGAATCTCCCTCTTTAGGACCTCTAGCAGCTAATATT 240
Db 1603 CCTCAGCAATGATGATGCGCTGGAATCTCCCTCTTTAGGACCTCTAGCAGCTAATATT 1662
Qy 241 TTTACTCTCTTTGGACCCCTGATCTTCAACTCTCTTTAAAGTTGCTCTTCCAGAAAT 300
Db 1663 GCTACTCTCTTTGGACCCCTGATCTTAACTCTCTTTAACTTTGCTCTTCCAGAAAT 1722
Qy 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAAACCCAGATGCAATGCAATGCTAA 360
Db 1723 CGAAGCTGTAAAGCTA-----CAAATGGAGCCCAAGATGCAATGCAATGCTAA 1770
Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGCTTAAATGACATGTA 420
Db 1771 GATCTACCGGAGACCCCTGGACCGGCTGCTAGCCACGATCTGATGTTAATGACATCAA 1830
Qy 421 AGTCACCCCTCCGAGGAAATCTCAACTGACACACCCCTACTACACTCCAAATTCAGTAGG 480
Db 1831 AGGCACCCCTCCTGAGGAATCTCAGCTGCACAACTCTACTACGCCCAATTCAGCAGG 1890
Qy 481 AAGCAGTTAGAGCAGTTGTGAGCAACCTCCCAACAGTAATCTTGGGTTTTCCTGTTGAGA 540
Db 1891 AAGCAGTTAGAGGCGTCTGGGCAACCTCCCAACAGCAGCTTAGGTTTTCCTGTTGAGA 1950
Qy 541 GGGTGGACTCAGACAGGAGTACTGGAATTTCTAGGCTGACTAGAAATCCCNAGACCT 600
Db 1951 TGGGGGACTCAGACAGGAGTACTGGAATTTCTAGGCTGACTAGAAATCCCTAAGCCT 2010
Qy 601 ANCTGGGAAGGTGACCCATCCATCTTTAAACATGCGGCTTGCAACTTACTCACAACCCG 660
Db 2011 AGCTGGGAAGGTGACCACTCCACCTTTAAACAGCGGCTTGCAACTTACTCACAACCTG 2070
Qy 661 ACCAATCAGAGAGTCTAATAATGCTAATCAGGCAAAAAACAGAGGTAAGCAATAGCC 720
Db 2071 ACCAATCAGAGAGTCTAATAATGCTAATTAGGCAAAAAACAGAGGTAAGCAATAGCC 2130
Qy 721 AATCATCTATTGCTGAGAGCAGCGGGAAGGAGGAGGATTCGGATATAAATCTCAGGCA 780
Db 2131 AATCATCTATTGCTGAGAGCAGCGGGAAGGAGGAGGATTCGGATATAAATCTCAGGCA 2190
Qy 781 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCACTTGTATGGGAGCTCTGTTTT 840
Db 2191 TTCGAGCGGCAACGCGCAACCCCTTTGGGTCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250
Qy 841 CACTCTATTTCACTCTATTAATCATGCAACTGCACTCTCTCTGGTCGGTGTGTTTTATGG 900
Db 2251 CATGCTATTTCACTCTATTAATCTTGAATCTGCAACTGCACTCTCTCTGGTCGATGTTTACGG 2310
Qy 901 CTCAAGCTGAGCTTTGCTCGCCATCCACCACTGCTGTTTGGCCACCGTCAACACCGCT 960
Db 2311 CTTGAGCTGAGCTTTGCTCGCCATCCACCACTGCTGTTTGGCCACCGTCAACACCGCT 2370
Qy 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTGCTGATCCAGCAGGCT 1020
Db 2371 GCTGACTTCCATCCCTCTGGATCATGCAAGGTTGTCGCTGTGCTGATCCAGCAGGCT 2430
Qy 1021 ACCAATGCACTCCCGATCAGGCTAAGGCTTGCATGTTCTCTGATGCTGCTAAGTGC 1080
Db 2431 ACCAATGCGGCTCCCAATTTGGGCTAAGGCTTGCATGTTCTCTGATGCTGCTAAGTGC 2490
Qy 1081 TGGGTTGCTCTAATGAAGCTGAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db 2491 TGGGTTGCTCTAATGAAGCTGAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2550

Qy 1141 CCACGGCTTCTAATAGAGCTATAAAGCTCACCAGCATGGCCCAAGATTCCATTCTTGGTA 1200
Db 2551 CCACAGCTTCTAATAGAGCTATAAAGCTCACCAGCATGGCCCAAGTTCCATTCTT-GAA 2609
Qy 1201 TCTGTAGGCCAAGAACCCAGGTGAGAGNAGTGGCTTGCACCAATTTGGGAG 1257
Db 2610 TCCATAAGGCCAAGAACCCAGGTGAGAGNAGTGGCTTGCACCAATTTGGGAG 2666

RESULT 14
AF520493S2 2694 bp DNA linear PRI 11-FEB-2004
LOCUS Homo sapiens individual 24 allele A, envelope glycoprotein gene,
DEFINITION complete cds, and 3' long terminal repeat, complete sequence.
AF520494
VERSION AF520494.1 GI:33410956
KEYWORDS
SOURCE 2 of 2
ORGANISM Homo sapiens (human)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2694)
AUTHORS Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G.,
Bonnaud, B., Lucotte, G., Duret, L. and Mandrand, B.
TITLE The endogenous retroviral locus ERVWE1 is a bona fide gene involved
in hominoid placental physiology
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
PUBMED 14757826
REFERENCE 2 (bases 1 to 2694)
AUTHORS Mallet, F., Bouton, O. and Oriol, G.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée
d'Italie, Lyon 69364 cedex 07, France

FEATURES
source
1..2694
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="individual 24 allele A"
/db_xref="taxon:9606"
/map="7q21-q22"
/sex="female"
/cell_type="PBMC's"
/note="Asian
endogenous_virus: HERV-W"
misc_feature
43..57
/note="splice acceptor site"
CDS
284..1900
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAQ17569.1"
/db_xref="GI:33410957"
misc_feature
1705..1720
/note="splice acceptor site"
1959..2694
ORIGIN
Query Match 77.9%; Score 1035.2; DB 9; Length 2694;
Best Local Similarity 90.1%; Pred. No. 3.1e-304;

Matches 1133; Conservative 0; Mismatches 111; Indels 13; Gaps 2;

QY 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAAACCGCAAAAGAGGGGGAACCTGTTTATTTTT 60
Db 1423 TCAAAATCGAAGAGCTTTAGACTTGTCTAAACCGCTGAAGAGGGGGAACCTGTTTATTTTT 1482

QY 61 AGGGGAGAGTCTGTGTAGTATCTTAATCAATCTGGAATCATTAATCTGAGAAAGTTAAGA 120
Db 1483 AGGGGAGAGTCTGTGTATTTATTTAATCAATCTGGAATCGTCACTGAGAAAGTTAAGA 1542

QY 121 AATTTGAGATCGAATAATAATGTAGAGCAGAGGACCTTCAAAAACATGTCACCTCTGGGGCCT 180
Db 1543 AATTCGAGATCGAATAACAGTAGAGCAGAGGAGCTTCGAAACACTGGACCTCTGGGGCCT 1602

QY 181 CCTCAGCAATGATGCCCTGGACTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATTT 240
Db 1603 CCTCAGCAATGATGCCCTGGATTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATTT 1662

QY 241 TTTACTCTCTTTGGACCCCTGTATCTTCAATCTTCTTGTAAAGTTGTCTCTTCCAGAAAT 300
Db 1663 GCTACTCTCTTTGGACCCCTGTATCTTTAACTCTTGTAACTTGTCTCTTCCAGAAAT 1722

QY 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAACCCAGATGCAGTCCATGACTAA 360
Db 1723 CGAAGCTGTAAACTA-----CAATGGAGCCCAAGATGCAGTCCAGACTAA 1770

QY 361 AATCTACCGTGGACCCCTGGACCGGCTGTCTAGACTATGCTCTGATGTTAATGACATTTGA 420
Db 1771 GATCTACCGCAGACCCCTGGACCGGCTGTCTAGCCACGATCTGATGTTAATGACATCAA 1830

QY 421 AGTCACCCCTCCGAGGAAATCTCACTGCAACACCCCTACTACATCTCAGATTCAGTAGG 480
Db 1831 AGGCACCCCTCTCAGGAAATCTCAGTGCACAACTCTACTACGCCCCCAATTCAGCAGG 1890

QY 481 AAGCAGTTAGCAGGTGTGAGCCCACTCCCAACAGTACTTGGGTTTCTGTTGAGA 540
Db 1891 AAGCAGTTAGCAGGTGTGAGCCCACTCCCAACAGCAGTACTTGGGTTTCTGTTGAGA 1950

QY 541 GGGTGAAGTCTGAGACAGGACTAGCTGATTTCTAGGCTGACTAAGAAATCCCAAGACCT 600
Db 1951 TGGGGGACTGAGACAGGACTAGCTGATTTCTAGGCTGACTAAGAAATCCCAAGACCT 2010

QY 601 ANCTGGAAGGTGACCGATCCATCTTTTAAACATGGGGCTTGCAACTTAGCTCACACCG 660
Db 2011 AGCTGGAAGGTGACCATCCACCTTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

QY 661 ACCAATCAGAGCTCACTAAATGCTAATCAGGCAAAACAGAGGTAAAGCAATAGCC 720
Db 2071 ACCAATCAGAGCTCACTAAATGCTAATTAGGCAAAACAGAGGTAAAGCAATAGCC 2130

QY 721 AATCATCTATTGCTGAGAGCAGACGGGGAAGGACAGGATTTGGGATATAAATCTCAGGCA 780
Db 2131 AATCATCTATTGCTGAGAGCAGACGGGGAAGGACATGATCGGATATAAATCCAGTC 2190

QY 781 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGATGGGAGCTCTGTTTT 840
Db 2191 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCTTTGATGGGAGCTCTGTTTT 2250

QY 841 CACTCTATTTCATCTTAAATCATGCAACTGACATCTTCTGCTCGTGTGTTTATGG 900
Db 2251 CATGCTATTTCACTCTATTAATCTTCAACTGACATCTTCTGCTCCATGTTTCTTACGG 2310

QY 901 CTCAGCTGAGCTTTTCTGCGCATCCACACTCTCTGTTTGGCCACCGTCAAGACCCGCT 960
Db 2311 CTTGAGCTGAGCTTTTCTGCGCATCCACACTCTCTGTTTGGCCACCGTCAAGACCCGCT 2370

QY 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCTGATCCAGCGAGGT 1020
Db 2371 GCTGACTTCCATCCCTCTGGATCATGACGGGTGTCCGCTGTGCTCTGATCCAGCGAGG 2430

QY 1021 ACCAATGCACTCCCGATCAGGCTAAAGGCTTGCCATTTGTTCTGATGCTTAAGTGCC 1080
Db 2431 ACCAATGCGGCTCCCAATCGGCTTAAAGGCTTGCCATTTGTTCTGATGCTTAAGTGCC 2490

QY 1081 TGGGTTTGTCTTAATAGAACTGAACACTGGTCACTGGGTTCATGGTTCCTCTTCCATGAC 1140
Db 2491 TGGGTTTCATCTTAATAGAACTGAACACTAGTCACTGGGTTCATGGTTCCTCTTGTGAC 2550

QY 1141 CCAAGGCTTCTTAATAGAGCTATAAACAACCTCACCGCATGGCCCAAGATTCCATTCTTGGTA 1200
Db 2551 CCACAGCTTCTTAATAGAGCTATAAACAACCTCACCGCATGGCCCAAGATTCCATTCTT-GAA 2609

QY 1201 TCTGTAGGCCCAAGAAACCCAGGTCAGAGAAAGTGTAGGCTTGCCACATTGGGAAG 1257
Db 2610 TCCATAAGGCCCAAGAAACCCAGGTCAGAGAAACAGAGGCTTGCCACCATCTTGGGAG 2666

RESULT 15
AF520495S2 2694 bp DNA linear PRI 11-FEB-2004
LOCUS
DEFINITION Homo sapiens individual 24 allele B, envelope glycoprotein gene, complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION AF520496
VERSION AF520496.1 GI:33410960
KEYWORDS
SEGMENT 2 of 2
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2694)
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2694)
Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G., Bonnaud, B., Lucotte, G., Duret, L. and Mandrand, B.
The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology
Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
14757826
REFERENCE 2 (bases 1 to 2694)
Mallet, F., Bouton, O. and Oriol, G.
Direct Submission
Submitted (07-JUN-2002) Retrovirology Department, UMR 2142 CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée d'Italie, Lyon 69364 cedex 07, France
FEATURES
source
1. 2694
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="individual 24 allele B"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-q22"
/sex="female"
/cell_type="PBMC's"
/notes="Asian
endogenous_virus: HERV-W"
misc_feature 43..57
/notes="splice acceptor site"
CDS 284..1900
/notes="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AA017570.1"
/db_xref="GI:33410961"
/translations="MALPHYIFLFTVLLPSFTLAPPCKMTSSSPYQBLWRMQRPNIDAPSYRSLSGKGTPTAHTMPNRCVHSATLCMHANTHYWTGKMINPSCGGLVGTCWYFTGTGMSDGGVQDQREKHVEVISQLTRVHGTSPPYKGLDLKJHETLRTHTLVLENTLTGLHEVSAQNPTNCWICLPLNFRPVSIPEQWNNFTEINTTSSALVGPLVSNLEIHTSNLTCTVKFSNTTYSQICIRWTPPTQIVCLPSGIFVCGTSAIYCLNGSSSMCFSLFVPPMTIYTQDLYSYVISPRNKRVPILFVIGAGVILGALGTGGTITSTQFYFKLSQELNGDMERVADSLVTQDLNLAALAVLNQRALDLTAE RGCTCLFGECCYVYQSGIVTEKVEIKRDRIQRABELRNTGPMGLLSQWMPWILP FLGPLAAIILLILFGPCIFNLVNFVSSRIEAVKLQWEPKQSKTKIYRPLDRPASPRSDNDIKGTPPEISAAQPLLRPNAGSS"
1705..1720
misc_feature
/notes="splice acceptor site"
LTR 1959..2694

ORIGIN

Query Match 77.9%; Score 1035.2; DB 9; Length 2694;
Best Local Similarity 90.1%; Pred. No. 3.1e-304;
Matches 1133; Conservative 0; Mismatches 111; Indels 13; Gaps 2;

QY 1 TCAAAATCGAAGAGCTTTAGACTGTTGCTAAACCGCCAAAGAGGGGGAACCTGTTTATTTT 60
Db TCAAAATCGAAGAGCTTTAGACTGTTGCTAAACCGCTGAAAGAGGGGGAACCTGTTTATTTT 1482

QY 61 AGGGGAAGATGCTGTTAGTATGTTAATCAATCTGGATCATTTACTGAGAAAGTTAAGA 120
Db AGGGGAAGATGCTGTTATTTATGTTAATCAATCTGGAAATCGTCACTGAGAAAGTTAAGA 1542

QY 121 AATTGAGATCGAATAAATGTTAGAGCAGAGGACCTTCAAAAACAATGCAACCTCGGGGCCT 180
Db AATTGAGATCGAATAAATGTTAGAGCAGAGGAGCTTCGAAAACCTGGACCTCGGGGCCT 1602

QY 181 CCTCAGCAATGAGTCCCTGGACTCTCCCTCTTTAGGACCTCTAGCAGCTATAATATT 240
Db CCTCAGCAATGAGTCCCTGGATCTCCCTCTTTAGGACCTCTAGCAGCTATAATATT 1662

QY 241 TTTACTCCTCTTTGGACCTGTTATCTCAACTTCCTTGTAACTTTGTCTCTTCCAGAA 300
Db GCTACTCCTCTTTGGACCTGTTATCTTTAACTCTCTTGTAACTTTGTCTCTTCCAGAA 1722

QY 301 TGAAGCTGTAAAGCTACAAATAGTTCTTTCAAATGGAAACCCAGATGCAGTCCATGACTAA 360
Db CGAAGCTGTAAACTA-----CAAATGGAGCCCAAGATGCAGTCCAGACTAA 1770

QY 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGATGTTAATGACATTTGA 420
Db GATCTACCGCAGACCCCTGGACCGGCTGTAGCCACGATCTGTGTAAATGACATCA 1830

QY 421 AGTCACCCCTCCGAGGAATCTCACTGCAACACCCCTACTACATCCGAATTCAGTAGG 480
Db AGGCACCCCTCCGAGGAATCTCAGCTGCAACACCTCTACTACGCCCCCAATTCAGCAG 1890

QY 481 AAGCAGTTAGAGAGTGTTCAGCAACCTCCCAACAGTACTTTGGGTTTTCCTGTTGAGA 540
Db AAGCAGTTAGAGGCTGTTCGGCAACCTCCCAACAGCACTTAGGTTTTCCTGTTGAGA 1950

QY 541 GGTGGAATGAGACAGGACTAGCTGGAATTCCTAGGCTGACTAAGAAATCCCAAGCCT 600
Db TGGGGGACTGAGACAGGACTAGCTGGAATTCCTAGGCTGACTAAGAAATCCCAAGCCT 2010

QY 601 ANCTGGGAAGTGACCGCTCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
Db AGCTGGGAAGGTGACCAATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070

QY 661 ACCAATCAGAGAGCTCACTAAATGCTTAATCAGGCAAAACAGAGGTTAAGCAATAGCC 720
Db ACCAATCAGAGAGCTCACTAAATGCTTAATGAGCAAGACAGAGGTTAAGAAATAGCC 2130

QY 721 AATCATCTATTGCTGAGACACAGCGGGAAGCAAGGATTTGGGATATAAACTCAGGCA 780
Db AATCATCTATTGCTGAGACACAGCAGGAGGACAATGATCGGATATAAACCCCAAGTC 2190

QY 781 TTCAAGCCAGCAACAGCAACCCCTTTGGTCCCTCCCTCCCATTTGATGGGAGCTCTGTTT 840
Db TTCAAGCCAGCAACAGCAACCCCTTTGGTCCCTCCCTCCCTTTGATGGGAGCTCTGTTT 2250

QY 841 CACTTATTTTCACTCTATTAAATCATGCAACTGCATCTTCTGCTCGCTGTTTATGG 900
Db CATGCTATTTCATCTATTAAATCTTGCAACTGCATCTTCTGCTCGATGTTTCTTACGG 2310

QY 901 CTCAAGCTGAGCTTTTGTTCGCCATCCACACTGCTGTTTGGCCACCGTCAACAGCCGCT 960
Db CTTGAGCTGAGCTTTGCTCGCCATCCACCCTGCTGTTTGGCCGACCGCAGACCCGCC 2370

QY 961 GCTGACTTCCATCCCTTTGATCCAGCAGAGTGTCACTGTGCTCTGATCCAGCAGGT 1020
Db GCTGACTTCCATCCCTCTGGATCATGCAGGGTGTCCGCTGTGCTCTGATCCAGCAGGC 2430

QY 1021 ACCATTGGCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCTCTGATGCTTAAGTGCC 1080
Db ACCATTGGCGCTCCCAATCGGCTAAAGGCTTGCCATTGTTCTCTGATGCTTAAGTGCC 2490

QY 1081 TGGGTTTGTCTTAATGAACTGAACACTGGTCACTGGGTTCCATGTTCTCTTCCATGAC 1140
Db TGGGTTTCACTCAATTAATGAGCTGAACACTAGTCACTGGGTTCCATGTTCTCTTCTGTGAC 2550

QY 1141 CCAGGCTTCTTAATAGAGCTATAAACAACCTCACCGCATGGCCCAAGATTCCATTCTTGGTA 1200
Db CCACGCTTCTTAATAGAGCTATAAACAACCTCACCGCATGGCCCAAGTTCCATTCTT-GAA 2609

QY 1201 TCTGTGAGGCGCAAGAACCCCGAGTCCAGAAANGTGGAGGCTTGCCACCATTTGGGAAG 1257
Db TCCATAAGGCCAAGAACCCCGAGTCCAGAGAACACGAGGCTTGCCACCATCTTGGGAG 2666

RESULT 16
AF520497S2 2694 bp DNA linear PRI 11-FEB-2004
LOCUS Homo sapiens individual 30 allele A, envelope glycoprotein gene,
DEFINITION complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION AF520498
VERSION AF520498.1 GI:33410964
SEGMENTS 2 of 2
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2694)
AUTHORS Mallet F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G.,
Bonnaud, B., Lucotte, G., Duret, L. and Mandrand, B.
TITLE The endogenous retroviral locus ERWE1 is a bona fide gene involved
JOURNAL in hominoid placental physiology
PUBMED Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
14757826
REFERENCE 2 (bases 1 to 2694)
AUTHORS Mallet F., Bouton, O. and Oriol, G.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée
d'Italie, Lyon 69364 cedex 07, France
FEATURES
1..2694
source /organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="individual 30 allele A"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-q22"
/sex="female"
/cell_type="PBMC's"
/note="Caucasian
endogenous_virus: HERV-W"
43..57
misc_feature /note="splice acceptor site"
284..1900
CDS /note="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAQ17571.1"
/db_xref="GI:33410965"
/translation="MALPYHIFLFTVLLPSFTLTAPPCRCMTSSPYQEFRLMRQRP
GNIDAPYKSLSGKPTPTFAHTHPMPCNTHSALCLMHANTHTWIKWNPSPCGGLGV
TVCTWYFTGTGMSDGGVDQARKEHVKEVISQLTRVHGTSSPYKGLDLSLUHETLRT
HPRVLSFNTLTTLGHEVSAQNTNCWICLPLNFRPYVSIPTVQWNNSTINTTSV
LVGPLVSLNEITHTNLTCVKPSNTTYTNSOCIRWTPPTQIVCLPSGIFVCGTSA
VRLGNSFSMCFLSNLPVPMPTIYTEQDLYSVISKPRNRPVILPFTVIGAGVIGALG
TGIGITSTQFYFKLSQELNGDMERVADSLVTLQDLSLAAVLQNERAIDLITAE
RGCTCLFGECECYTVNQSGIVTEKVKERIDRIQRABELRNTGPWLLSQWNPWILP
FLGPLAAIILLFGLFCIFNLLNFVSSRIEAVKLQMEPKMQSKTKIYRRLDRPASP

```

misc_feature
LTR
ORIGIN
Query Match
Best Local Similarity 77.9%; Score 1035.2; DB 9; Length 2694;
Matches 1133; Conservative 0; Mismatches 111; Indels 13; Gaps 2;
1 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCAAGAGAGGGGGAACCTGTTTATTTT 60
1423 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCTGAAGAGGGGGAACCTGTTTATTTT 1482
61 AGGGGAAGAAATGCTGTGTATGTTAATCAATCTGGAATCATPACTGAGAAAGTTAAAGA 120
1483 AGGGGAAGAAATGCTGTGTATGTTAATCAATCTGGAATCATPACTGAGAAAGTTAAAGA 1542
121 AATTTGAGATCGAATATATGATAGAGAGGAGGACCTTCAAAACATGCAACCTGGGCGCT 180
1543 AATTCGAGATCGAATATCAACGATAGAGAGGAGGAGCTTCAAAACATGCAACCTGGGCGCT 1602
181 CTTACGCCAATGATGATCCCTGGAGCTCTCCCTCTTTAGGACCTCTAGCAGCTATATATT 240
1603 CTTACGCCAATGATGATCCCTGGAGCTCTCCCTCTTTAGGACCTCTAGCAGCTATATATT 1662
241 TTTACTCTCTTTTGGACCTCTATCTCAATCTTCTTGTAAAGTTGTCTCTTCCAGAAAT 300
1663 GCTACTCTCTTTGGACCTCTATCTTAACTCTCTTGTAACTTGTCTCTTCCAGAAAT 1722
301 TGAAGCTGTAAAGCTCAAAATGATCTTCAATGGAACCCAGATCGAGTCAATGACTAA 360
1723 CGAAGCTGTAAAGCTCAAAATGATCTTCAATGGAACCCAGATCGAGTCAAAAGCTAA 1770
361 AATCTACCTGGAGACCTCTGACCGGCTGTAGACTATGCTCTGATGTTGTTAAAGACATGA 420
1771 GATCTACCGAGACCTCTGACCGGCTGTAGCCACGATCTGATGTTAAAGACATCAA 1830
421 AGTCACCTCTCCGAGGAAATCTCAATGCAACACCCCTACTACACTCAATTCAGTAGG 480
1831 AGGCACCTCTCTGAGGAAATCTCAGCTGCACAACTCTACTACGCCCAATTCAGCAGG 1890
481 AGCAGTTAGAGAGCTTGTAGCAGCACTCTCCCAACAGTACTTGGGTTTCTGTTGAGA 540
1891 AAGCAGTTAGAGGCTGCTGGGCAACTCTCCCAACAGCAGTCTAGGTTTCTGTTGAGA 1950
541 GGTGAGCTGAGAGAGAGTACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCNAAAGCT 600
1951 TGGGGGAGCTGAGAGAGAGTACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCNAAAGCT 2010
601 ANCTGGGAAGGTGACCGCATCTCAATCTTTAAACATGGGGCTTGCAACTTACTACACCCG 660
2011 AGCTGGGAAGGTGACCGCATCTCAATCTTTAAACATGGGGCTTGCAACTTACTACACCTG 2070
661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
2071 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 2130
721 AATCATCTATTGCTGAGAGCAGCGGAGGAGGACAGGATTCGGGATATAAATCTCAGGCA 780
2131 AATCATCTATTGCTGAGAGCAGCGGAGGAGGACAGGATTCGGGATATAAATCTCAGGCA 2190
781 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCTCCCTCCCATTTGATGGGAGCTCTGTTT 840
2191 TTGAGCCGCAACAGCAACCCCTTTGGGTCCTCCCTCCCATTTGATGGGAGCTCTGTTT 2250
841 CACTCTATTCTCTATTAAATCATGCAATGCACTCTTCTGCTGGTCCGTTTATGAG 900
2251 CATGCTATTCTCTATTAAATCTTGAACATGCACTCTTCTGCTGGTCCGTTTATGAG 2310
901 CTCAGCTGAGCTTTGTTGCGCATCCACACCTGCTGTTTGGCCACCGTCAACAGCCGCT 960
2311 CTTGAGCTGAGCTTTGCTGCGCATCCACCACTGCTGTTTGGCCGCGCCACCGCAGCCGCT 2370
```

```

961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCTGATCCAGCGAGT 1020
2371 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCTGATCCAGCGAGC 2430
1021 ACCATTGGCACTCCCGATCAGGCTAAAGGCTTGCATTGTTCTCTGATCCAGTAAAGTGC 1080
2431 ACCATTGGCACTCCCGATCAGGCTAAAGGCTTGCATTGTTCTCTGATCCAGTAAAGTGC 2490
1081 TGGGTTTGTCTTAATAGAACTCAACACTGCTCACTGGGTTCCATGTTCTCTTCCATGAC 1140
2491 TGGGTTTGTCTTAATAGAACTCAACACTGCTCACTGGGTTCCATGTTCTCTTCTGTGAC 2550
1141 CCACGGCTTCTTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCATTCCTTGGTA 1200
2551 CCACGGCTTCTTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCATTCCTTGGTA 2609
1201 TCTGTGAGGCCAAGAACCCAGCTCAGAGAAAGTGTGAGGCTTGCACCATTTGGGAAG 1257
2610 TCTGTGAGGCCAAGAACCCAGCTCAGAGAAAGTGTGAGGCTTGCACCATTTGGGAAG 2666

RESULT 17
AF520503S2 2694 bp DNA linear PRI 11-FEB-2004
LOCUS
DEFINITION
Homo sapiens individual 34 allele B, envelope glycoprotein gene,
complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION
AF520504
VERSION
AF520504.1 GI:33410976
KEYWORDS
2 of 2
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2694)
AUTHORS
Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
The endogenous retroviral locus ERVWE1 is a bona fide gene involved
in hominoid placental physiology
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
PUBMED
14757826
REFERENCE
2 (bases 1 to 2694)
AUTHORS
Mallet,F., Bouton,O. and Oriol,G.
Direct Submission
TITLE
Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée
d'Italie, Lyon 69364 cedex 07, France
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
PUBMED
14757826
FEATURES
source
1..2694
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="individual 34 allele B"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-q22"
/sex="male"
/cell_type="PBMC's"
/note="endogenous_virus: HERV-W"
misc_feature
43..57
/note="splice acceptor site"
284..1900
/note="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAQ17574.1"
/db_xref="GI:33410977"
/translation="MALPYHIFLTVLLPSFLTAPPCCRCMTSSPYQEFLLWRMRP
GNIDAPSVRSLSKGTPTTAHTHPRNCYHSATLCMHANTHYMTGMINPSCPGILV
TVCWTYFTGTGSDGGVQDQAREKHVKEVISLTVHGTSSPYKGLDLSKLHETLRT
HTRVLSLENTLTGLHEVSAONPTNCWICLPLNFRPVSI PVPEONNFSLEINTTSV
LVGPLVSNLEIHTSNLTCKFSNTYTTNSQICIRWVTPPTQIVCLPSGIFVCGTSA
YRCLNGSSSESMCFSLFSLVPPMTIYTEQDLISYVISKPRNKRVPFLFPVIGAGVIGALG
```

Query Match 77.9%; Score 1035.2; DB 9; Length 2694;
Best Local Similarity 90.18; Pred. No. 3.1e-304;
Matches 1133; Conservative 0; Mismatches 111; Indels 13; Gaps 2;
ORIGIN
TGIGGTTSTQFYFKLSOELNGDMERVADSLVTLQDQLNSLAIAVVLQNRALDLTAE
RGGTCFLGSECCYYNQSGIVTEKVKIEIRDRIORRABELRNTGPNGLLSQWMPWILP
FLGPLAAIILLLFLGFCIFNLVNFVSRILEAVKLQWEPKQSKTKIYRRLDRPASP
RSDVNDIKGTPPPEISAAQPLLRPNAGSS"
1705..1720
misc_feature
/note="splice acceptor site"
1959..2694

Qy 1 TCAAAATCGAAGAGCTTTAGACTTGTCAACCGCCAAAGAGGGGGAACCTGTTATTTT 60
Db 1423 TCAAAATCGAAGAGCTTTAGACTTGTCAACCGCTGAAAGAGGGGGAACCTGTTATTTT 1482

Qy 61 AGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTTACTGAGAAAGTTAAGA 120
Db 1483 AGGGAAGAATGCTGTTATTTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAGA 1542

Qy 121 AATTTGAGATCGAATAAATGTTAGAGCAGAGGACCTTCAAAACACTGCACCCCTGGGSCCT 180
Db 1543 AATTCGAGATCGAATACAGCTAGAGCAGAGGAGCTTTCGAAACACTGGACCCCTGGGSCCT 1602

Qy 181 CCTCAGCAATGAGTCCCTGGACTCTCCCTCTTTTAGGACCTCTAGCAGCTATAATATT 240
Db 1603 CCTCAGCAATGAGTCCCTGGACTCTCCCTCTTTTAGGACCTCTAGCAGCTATAATATT 1662

Qy 241 TTTACTCTCTTTGGACCCCTGTATCTCAACTCTTCTTTGTTAAGTTTGTCTCTTCCAGAA 300
Db 1663 GCTACTCTCTTTGGACCCCTGTATCTTTAACTCTCTTTGTTAAGTTTGTCTCTTCCAGAA 1722

Qy 301 TGAAGCTGTAAGCTACAATAGTTCTTCAAAATGGAACCCAGATGCGAGTCCATGACTAA 360
Db 1723 CGAAGCTGTAAGCTA-----CAAAATGGAACCCAGATGCGAGTCCAGACTAA 1770

Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGTTAATGACATGTA 420
Db 1771 GATCTACCGCAGACCCCTGGACCGGCTGCTAGCCACGATCTGATGTTAATGACATCA 1830

Qy 421 AGTCACCCCTCCGAGGAATCTCACTGCAACCCCTACTACATCCGAAATTCAGTAGG 480
Db 1831 AGGCACCCCTCCTGAGGAATCTCAGCTGCAACACCTCTACTAGCCCAATTCAGCAGG 1890

Qy 481 AAGCAGTTAGCAGAGTTGTGAGCAACCTCCCAACAGTACTTGGGTTTCTGTTGAGA 540
Db 1891 AAGCAGTTAGCAGGCTGCTGGCAACCTCCCAACAGCAGCTTAGGTTTCTGTTGAGA 1950

Qy 541 GGGTGGACTGAGAGACAGGACTAGCTGGAATTCCTAGGCTGACTAAGAAATCCCAAGCCT 600
Db 1951 TGGGGACTGAGAGACAGGACTAGCTGGAATTCCTAGGCTGACTAAGAAATCCCTAAGCCT 2010

Qy 601 ANCTGGGAAGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTAGCTCACACCG 660
Db 2011 AGCTGGGAAGTGACCAATCCACCTTTAAACACGGGGCTTGCAACTAGCTCACACCTG 2070

Qy 661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGCAAAAACAGAGTAAAGCAATAGCC 720
Db 2071 ACCAATCAGAGAGCTCACTAAATGCTAATAGCAAGACAGAGTAAAGAAATAGCC 2130

Qy 721 AATCATCTATTGCTGAGAGCAGCGGGAAGGACAGGATTTGGGATATAAATCAGGCA 780
Db 2131 AATCATCTATTGCTGAGAGCAGCAGGAGGAGCAATGATCGGATATAAACCAGTC 2190

Qy 781 TTCAAGCAGCAACAGCAACCCCTTTGGTCCCTCCCAATTTGATGGGAGCTCTGTTT 840
Db 2191 TTCGAGCGGCAACGGCAACCCCTTTGGTCCCTCCCTTTGATGGGAGCTCTGTTT 2250

Qy 841 CACTCTATTTCACTCTATTAAATCATGCACTGCACTCTTCTGCTCGCTGTTTTTATGG 900
Db 2251 CATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGCTCCTGTTCTTACGG 2310

Qy 901 CTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCACCGTCAACACCGGCT 960
Db 2311 CTTCAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGCACCGTCAACACCGGCT 2370

Qy 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCTGATCCAGCGAGGT 1020
Db 2371 GCTGACTTCCATCCCTTCTGATCATCAGGAGTGTCCGCTGTCTCTGATCCAGCGAGG 2430

Qy 1021 ACCATTGGCAGCTCCCGATCAGGCTAAAGGCTTGCATTTGCTTCCATGCTGAAGTGCC 1080
Db 2431 ACCATTGGCAGCTCCCGATCAGGCTAAAGGCTTGCATTTGCTTCCATGCTGAAGTGCC 2490

Qy 1081 TGGTTTGTCTTAATAGAACTGAACACTGCTGCTCACTGGGTTCCATGTTCTTCCATGAC 1140
Db 2491 TGGTTTCACTCAATTAATGAGCTGAACACTAGTCACTGGGTTCCATGTTCTTCTCTGAC 2550

Qy 1141 CAAGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCAGATTCATTTCTTGGTA 1200
Db 2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCAGGTTCCATTTCTT-GAA 2609

Qy 1201 TCTGTGAGGCGCAAGAACCCAGCTCAGAAAGTGGAGGTTGCCACCATTTGGGAAG 1257
Db 2610 TCATTAAGCCCAAGAACCCAGCTCAGAGAACACAGAGGCTTGCACCATTTCTGGAG 2666

RESULT 18
AF520509S2 2694 bp DNA linear PRI 11-FEB-2004
LOCUS
DEFINITION
Homo sapiens individual 45 allele A, envelope glycoprotein gene,
complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION
VERSION
KEYWORDS
SEGMENT
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1..2694
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="individual 45 allele A"
/db_xref="taxon:9606"
/chromosomes="7"
/map="7q21-q22"
/sex="male"
/cell_type="PBMC's"
/note="endogenous_virus: HERV-W"
43..57
/note="splice acceptor site"
misc_feature
284..1900
/note="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAQ17577.1"
/db_xref="GI:33410989"
/translation="MALPVIHFLFTVLLPSFTLTAPPKRCWMTSSSPYQEFLLRMORP
GNIDAPYSLSKGTPTFTTAHTMPNCHSATLCHNHTWTKMNPSCPGGLGV
TWCYFTTGTGMSDGGVQDQAREKHKEVISQLTRKHVHTSSPYKGLDLSKLHETLRT

```
HTRLVLENTLLTGLHEVSAONPTNCWICLPLNFRPVVSIPIVPEOWNFSTEINTTSSV
LVGLPLVNLLEIHTSNLTCVKEFNTTYTNSOCIRWVTPPTQIOLVCLPSGIFVFCGSA
YRLCNGSSSMCFSLFVPMYIYTEDQDIYSYVISKRNKRVFILLFVIGAGVGLGALG
TGIGGITTSTQFYKLSQBELNGDMERVADSLVTLQDOLNSLAAVVLONRRADLLTAE
RGCTCLFGEBCCYNQSGIVTEKVEIRDIQRABEILRNTPWGLLSOMWPTLP
FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQWPMQSKTKIYRRLDRPASP
RSDVNDIKGTPPPEISAAQPLLRPNSSGS"
misc_feature
1705..1720
/notes="splice acceptor site"
1959..2694
ORIGIN
Query Match 77.9%; Score 1035.2; DB 9; Length 2694;
Best Local Similarity 90.1%; Pred. No. 3.1e-304; Indels 13; Gaps 2;
Matches 1133; Conservative 0; Mismatches 111;
QY 1 TCAAAATCGAAGAGCTTTAGACTTGTCAACCGCCAAAGAGGGGGAACTGTTTATTTT 60
DB 1423 TCAAAATCGAAGAGCTTTAGACTTGTCAACCGCTGAAGAGGGGGAACTGTTTATTTT 1482
QY 61 AGGGGAAGAAATGCTGTGTAGTATGTTTAAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
DB 1483 AGGGGAAGAAATGCTGTGTATGTTTAAATCAATCGGAATCGTCACTGAGAAAGTTAAAGA 1542
QY 121 AATTTGAGATCGAATATAATGTAGACAGAGGACCTTCAAAACACTGCACCCCTGGGGCCT 180
DB 1543 AATTTGAGATCGAATACAACGTAGACAGAGGAGCTTCGAAACACTGGACCCCTGGGGCCT 1602
QY 181 CCTCAGCAATGATGATGCCCTGGGACTCTCCCTCTTTAGGACCTCTAGCAGCTATAATATT 240
DB 1603 CCTCAGCAATGATGATGCCCTGGGATCTCCCTCTTTAGGACCTCTAGCAGCTATAATATT 1662
QY 241 TTTTACTCTCTTTGGACCCCTGTATCTTCAACTCTTCTGTGTAAAGTTGTCTCTTCCAGAAT 300
DB 1663 GCTACTCTCTTTGGACCCCTGTATCTTAACTCTTGTGTAACTTGTCTCTTCCAGAAT 1722
QY 301 TGAAGCTGTAAAGCTACAATAGTTCTTCAAATGGAAACCCAGATGCAAGTCCATGACTAA 360
DB 1723 CGAAGCTGTAAAGCTA-----CAAATGGAGCCCAAGATGCAAGTCCCAAGACTAA 1770
QY 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGATGTGTTAATGACATTTGA 420
DB 1771 GATCTACCGAGACCCCTGGACCGGCTGTAGCCACGATCTGATGTGTTAATGACATCAA 1830
QY 421 AGTCACCCCTCCGAGGAAATCTCAACTGCACAAACCCCTACTACACTTCAATTCAGTAGG 480
DB 1831 AGGCACCCCTCCTGAGGAATCTCAGCTGCACAACTCTACTACGCCCAATTCAGCAGG 1890
QY 481 AAGCAGTTAGAGAGTTGTGAGCAACCTCCCAACAGTACTCTGGGTTTCTGTTGAGA 540
DB 1891 AAGCAGTTAGAGCGGTGCTCGGCCAACCTCCCAACAGCACTTAGGTGTTTCTGTTGAGA 1950
QY 541 GGGTGGACTCAGAGACAGGACTAGTGGATTTCTTAGGCTGACTAAGAAATCCCNAGCCT 600
DB 1951 TGGGGGACTCAGAGACAGGACTAGTGGATTTCTTAGGCTGACTAAGAAATCCCTAAGCCT 2010
QY 601 ANCTGGGAAGTGACCCGCTCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCCG 660
DB 2011 AGCTGGGAAGTGACCCATCTCCCTTTAAACAGCGGGCTTGCAACTTAGCTCACACCTG 2070
QY 661 ACCAATCAGAGAGCTCACTAAATGTCTAATCAGGCAAAAAACAGAGGTAAAGCAATAGCC 720
DB 2071 ACCAATCAGAGAGCTCACTAAATGTCTAATTAGGCAAAACAGAGGTAAAGAAATAGCC 2130
QY 721 AATCATCTATTGCTTGAGACACAGCGGGAAGGACAGGATTCGGATATAAATCTCAGGCA 780
DB 2131 AATCATCTATTGCTTGAGACACAGGAGGAGGACAAATGATCGGATATAAATCCCAAGTC 2190
QY 781 TTCAAGCGCAGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGTATGGGAGCTCTGTTTT 840
DB 2191 TTGAGCGCGCAACGGCAACCCCTTTGGGTCCCTCCCTTTGTATGGGAGCTCTGTTTT 2250
QY 841 CACTCTATTTTCACTCTATTAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTATGG 900
```

```
DB 2251 CATGCTATTTCACCTCTATTAAATCTTGCACATGCACTCTTCTCGTCAATGTTTCTTACGG 2310
QY 901 CTCAACTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGGCCACCGTCAACAGACCGCT 960
DB 2311 CTTGAGCTGAGCTTTTCGCTCGCCATCCACCACTGCTGTTTGGCCGCCACCGCAGACCGCC 2370
QY 961 GCTGACTTCCATCCCTTTGGATCCAGAGAGTGTCCACTGTGCTCTGTATCCAGCGAGGT 1020
DB 2371 GCTGACTCCCATCCCTCTCGATCATGAGGGTGTCCGCTGTCTCTGATCCAGCGAGGC 2430
QY 1021 ACCATTGGCACTCCCGATCAGGCTAAAGCTTGCATTTGCTCTGATGGTCAAGTGCC 1080
DB 2431 ACCATTGGCGTCCCAATCGGGCTTAAGGCTTGCCATTTGCTCTGATGGCTAAGTGCC 2490
QY 1081 TGGGTTTGTCTTAATAGAACTGAACACTGGTCTCACTGGGTTCCATGTTCTTTCATGAC 1140
DB 2491 TGGGTTTCACTTAATGAGCTGAACACTAGTCACTGGGTTCCATGTTCTTCTCTGTGAC 2550
QY 1141 CACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCATTTCTTTGGTA 1200
DB 2551 CCACAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTTCTTT-GAA 2609
QY 1201 TCTGTGAGCCCAAGAACCCAGCTCAGAGAAAGTGGAGGTTGCCACCATTTGGGAAG 1257
DB 2610 TCCATAAGCCCAAGAACCCAGGTCAGAGAAACAGAGGCTTGCCACCATCTTGGGAG 2666

RESULT 19
AF52051382 2694 bp DNA linear PRI 11-FEB-2004
LOCUS Homo sapiens individual 49 allele A, envelope glycoprotein gene,
DEFINITION complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION AF520514
VERSION AF520514.1 GI:33410996
KEYWORDS
SEGMENT 2 of 2
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2694)
AUTHORS Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
TITLE The endogenous retroviral locus ERYWE1 is a bona fide gene involved
in hominoid placental physiology
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
PUBMED 14757826
REFERENCE 2 (bases 1 to 2694)
AUTHORS Mallet,F., Bouton,O. and Oriol,G.
DIRECT SUBMISSION Direct Submission
TITLE Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
JOURNAL CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée
d'Italie, Lyon 69364 cedex 07, France
FEATURES
source
1..2694
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="individual 49 allele A"
/db_xref="taxon:9606"
/map="7q21-q22"
/sex="female"
/chromosomes="7"
/cell_type="PBMC's"
/notes="Asian
endogenous_virus: HERV-W"
misc_feature 43..57
/notes="splice acceptor site"
CDS 284..1900
/notes="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAQ17579.1"
```



```
/note="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAQ17580.1"
/db_xref="GI:33411001"
/translations="MALPVIHIFLFTLLPSFTLTAPPICRCMTSSPYOEFILWRMORP
GNIDAPSYRLSGTFTTAHTMPNRYHSTLQWANTHYWTGMWNPSPCGGLGV
TUCWTYPTQGSDDGGVQDOAREKHVKSIVLSQTHVGTSSPKGLDLISKLHETLR
HTRVLSFNLTLLGLHEVSANFNCTNWCILPLNFRPYVSI PVEPQNNSTEINTISV
LVGLVSNLEIBTHSLNLTCKFNTNTYNSQCIRWVTPPTQVILPSPGIFVCGTSA
YRCLGSSSMCSFLSKLSPMTIYEQDLYSYVISKPRNKRPVILPFI VAGVIGALG
TGTGGITSTQFYKLSQLBQMDVADSLVTLQDLNSLAALVQLNRRALDLITAE
RGCTCLFLGECVCYVNSCIVTEKVEIRDRIOREABELRNTGPMGLLSOMWPTLP
FLGPLAAILLLLFPICINLLVNFVSSRIEAVKLQWEPKMQSKTIYRPLDRPASP
RSDVNDKIGIPPEISAAQPLLRNSAGS"
misc_feature
1705..1720
/notes="splice acceptor site"
1959..2694
LTR
ORIGIN
Query Match 77.9%; Score 1035.2; DB 9; Length 2694;
Best Local Similarity 90.4%; Pred. No. 3; le-304;
Matches 1133; Conservative 0; Mismatches 111; Indels 13; Gaps 2;
1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAGAGGGGGACCTGTTTATTTT 60
1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAGAGGGGGACCTGTTTATTTT 1482
61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTTACTGAGAAAGTTAAAGA 120
1483 AGGGGAAGAATGCTGTTATGTTAATCAATCTGGAATCGTCACTGAGAAAGTTAAAGA 1542
121 AATTTGAGATCGAATAATAATGTTAGAGCAGAGGACCTTCAAAAACACTGCACCTCGGGCCCT 180
1543 AATTCGAGATCGAATAACAAGTAGAGCAGAGGAGCTTGGAAACACTGGACCTCGGGCCCT 1602
181 CCTCAGCAATGAGTCCCTGGAGCTCTCCCTCTTCTTAGGACCTCTAGCAGCTATAATTT 240
1603 CCTCAGCAATGAGTCCCTGGAGCTCTCCCTCTTCTTAGGACCTCTAGCAGCTATAATTT 1662
241 TTTACTCTCTTTGGACCTGTATCTCAACTCTCTTGTAAAGTTGTCTCTTCCAGAAAT 300
1663 GTACTCTCTTTGGACCTGTATCTTAACCTCTTGTAACTTTGTCTCTTCCAGAAAT 1722
301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAAACCCAGATGCAGTCCATGACTAA 360
1723 CGAAGCTGTAAACTA-----CAAATGGAGCCCAAGATGCAGTCCAGACTAA 1770
361 AATCTACCGTGGACCCCTGGACCGGCTGTCTAGACTATGCTCTGATGTTAATGACATGTA 420
1771 GATCTACCGCAGACCCCTGGACCGGCTGTCTAGCCACGATCTGATGTTAATGACATCAA 1830
421 AGTCACCCCTCCGAGGAATCTCACTGACACACCCCTACTACACTCCCAATTCAGTAGG 480
1831 AGGCACCCCTCCGAGGAATCTCAGTGCACACACCTCTACTACGCCCCCAATTCAGCAGG 1890
481 AAGCAGTTAGACGAGTTGTGAGCCAACTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 540
1891 AAGCAGTTAGACGAGTTGTGAGCCAACTCCCAACAGCAGTACTGTTTCTGTTGAGA 1950
541 GGGTGGACTGAGACAGCAGTACGTGGAATTTCTAGGCTGACTAAGAAATCCCNAAAGCCT 600
1951 TGGGGGACTGAGACAGCAGTACGTGGAATTTCTAGGCTGACTAAGAAATCCCTAAGCCT 2010
601 ANCTGGGAAGTGACCCGATCCATCTTTAAACATGGGGCTTGCAACTAGCTACACCCG 660
2011 AGCTGGGAAGTGACCAATCCATCCACTTTAAACACAGGGGCTTGCAACTAGCTACACCTG 2070
661 ACCAATCAGAGCTCACTAAATGCTAATCAGSCAAACACAGGAGTAAAGCAATAGCC 720
2071 ACCAATCAGAGCTCACTAAATGCTAATCAGSCAAACACAGGAGTAAAGCAATAGCC 2130
721 AATCATCTATTGCTGAGAGCAGACGCGGGAAGGACAGGATTTGGGATATAAACTCAGGCA 780
```

```
|||||
2131 AATCATCTATTGCTGAGAGCAGACGAGGAGGACAATGATCGGATATAAAACCAAGTC 2190
781 TTCAAGCCAGCAACAGCAACCCCTTTGGGTGCCCTTCCCATTTGATGAGGAGCTCTGTTTT 840
2191 TTGAGCCGCGCAACGCGCAACCCCTTTGGGTGCCCTTCCCATTTGATGAGGAGCTCTGTTTT 2250
841 CACTCTATTCTACTCTATTAAATCATGCAACTGCACCTCTTCTGGTCCGCTGTTTTTATGG 900
2251 CATGCTATTCTACTCTATTAAATCTTTGCAACTGCACCTCTTCTGGTCCATGTTTCTTACGG 2310
901 CTCAACTCAGAGCTTTGTTTCGCGCATCCACCACTGCTGTTTGGCCACCGTCACAGACCCGCT 960
2311 CTTGAGCTGAGCTTTGCTCGCATCCACCACTGCTGTTTGGCCACCGTCACAGACCCGCT 2370
961 GTGACTTTCATCCCTTTTGGATCCAGCAGAGTGTCCACTGTGCTCTGATCCAGCGAGGT 1020
2371 GCTGACTCCCATCCCTCTCGATCATGAGGCTGTCCGCTGTCTCTGATCCAGCGAGGC 2430
1021 ACCATTGCACTCCCGATCAGCTAAAGCTTGCATTTGCTCTCATGGCTTAAGTGCC 1080
2431 ACCATTGCGCTTCCCAATCGGGCTAAAGCTTGCATTTGCTCTCATGGCTTAAGTGCC 2490
1081 TGGGTTGCTCTAATAAGAACTGAACTGCTGCTCACTGGGTTTCCATGCTCTCTTCCATGAC 1140
2491 TGGGTTCACTCTAATTGAGCTGAACTAGTCACTAGTCACTGGGTTTCCATGCTCTCTTCTGAC 2550
1141 CCACGCTTCTAATAGAGCTATAAACACTCACCGCATGGCCCAAGATTCCATCTCTTGGTA 1200
2551 CCACGCTTCTAATAGAGCTATAAACACTCACCGCATGGCCCAAGATTCCATCTCTT-GAA 2609
1201 TCTGTAGGCGCAAGAACCCAGGTGAGAAAGTGGCTTGCACCACTTTGGGAAG 1257
2610 TCCATAAGGCCAAGAACCCAGGTGAGAAACAGAGGCTTGCACCACTTTGGGAG 2666
```

RESULT 21
AF520517S2 2694 bp DNA linear PRI 11-FEB-2004
Homo sapiens individual 55 allele A, envelope glycoprotein gene,
complete cds, and 3' long terminal repeat, complete sequence.
AF520518
AF520518.1 GI:33411004

2 of 2
Homo sapiens (human)
Homo sapiens
Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2694)
Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G.,
Bonnaud, B., Lucotte, G., Duret, L., and Mandrand, B.
The endogenous retroviral locus ERVWE1 is a bona fide gene involved
in hominoid placental physiology
Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
14757826
2 (bases 1 to 2694)
Mallet, F., Bouton, O. and Oriol, G.
Direct Submission
Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée
d'Italie, Lyon 69364 cedex 07, France
Location/Qualifiers
1..2694
/organism="Homo sapiens"
/mol_type="Genomic DNA"
/isolates="individual 55 allele A"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-q22"
/sex="female"
/cell_type="PBMC's"
/note="endogenous_virus: HERV-W"

source
1..2694

```
misc_feature 43..57
/notes="splice acceptor site"
CDS
284..1900
/notes="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAQ17581.1"
/db_xref="GI:33411005"
/translation="MALPHYFLFTVLLPSFTLTAPPKRCMTSSSPYQBEFLWRMQR
GNIDAPSYSLSGTFTFAHTHPRNCVHSATLCMHANTHYWTGMINPSCPGLGV
TVCWTFYTSLSGGGVOQAREKHVEVISOLTRVHGTSYPYKGLDLISKLHETURT
HTELVSLETNTLGLHVSQNPNCWICLPLNFRPVSI PVPEQNNFSEINTTSV
LVGPLVSNLEITHSNLTCVKFSNTTYTNSQCIRWPTQIIVCLPSGIFVFCGISA
YRLGSSSEMSCLFSLVPPMTIYTBODLYNVVISKPRNRPVILFVIGAGVIGALG
TGIGTSTYTFYKLSQELNGDMRVDSLVTLOQLNSLAALVILNRRALDILTAE
RGCTCLFLEGECYVYNQSIGVTEKVEIRDRIOREAEIRNTPGWLISQWMPWILP
FLGPLAAIILLILFGPCIFNLVNFVSSRIEAVKLQWEPKMQSKTKIYRPLDRPAS
RSDVNDIKTPPEISIAQPLLRPSAGSS"
misc_feature
1705..1720
/notes="splice acceptor site"
1959..2694
ORIGIN
77.9%; Score 1035.2; DB 9; Length 2694;
Query Match
Best Local Similarity 90.1%; Pred. No. 3.1e-304;
Matches 1133; Conservative 0; Mismatches 111; Indels 13; Gaps 2;
QY 1 TCAAAATCGAAGAGCTTAGACTTGTACCGGCAAAAGAGGGGGAACCTGTTTATTTT 60
DB 1423 TCAAAATCGAAGAGCTTAGACTTGTACCGGCAAAAGAGGGGGAACCTGTTTATTTT 1482
QY 61 AGGGGAAGATGCTGTATGTTATCAATCTGAATCTGAATCAATCTGAAGAACTTAAAGA 120
DB 1483 AGGGGAAGATGCTGTATGTTATCAATCTGAATCTGAATCAATCTGAAGAACTTAAAGA 1542
QY 121 AATTGAGATCGAATATAATCTAGACGAGGACCTTCAAAACACTGCACCTGGGGCCT 180
DB 1543 AATTGAGATCGAATATAATCTAGACGAGGACCTTCAAAACACTGCACCTGGGGCCT 1602
QY 181 CCTCAGCAATGATGCGCTGGACTCTCCCTCTCTTAGGACCTCTAGCAGCTATAATTT 240
DB 1603 CCTCAGCAATGATGCGCTGGACTCTCCCTCTCTTAGGACCTCTAGCAGCTATAATTT 1662
QY 241 TTACTCTCTTTGGACCTGTATCTTCAATCTCTTGTGTAAAGTTGTCTCTTCCAGAA 300
DB 1663 GCTACTCTCTTTGGACCTGTATCTTCAATCTCTTGTGTAAAGTTGTCTCTTCCAGAA 1722
QY 301 TGAAGCTGTAAGCTACAATAGTTCTTCAATGGAACCCAGACAGTCCAGTCACTAA 360
DB 1723 CGAAGCTGTAAGCTACAATAGTTCTTCAATGGAACCCAGACAGTCCAGTCACTAA 1770
QY 361 AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGTTAATCACTTGA 420
DB 1771 GATCTACCGGACCCCTGGACCGGCTGCTAGCCACGATCTGATGTTAATCACTTGA 1830
QY 421 AGTCACCCCTCCGAGGAAATCTCAATGCAACACCCCTACTACACTCCAATTCAGTAGG 480
DB 1831 AGGCACCCCTCCGAGGAAATCTCAGCTGCAACACCTCTACTACGCCCAATTCAGCAG 1890
QY 481 AAGCAGTTAGCAGGTTGTGAGCAACCTCCCAACAGTACTTGGGTTTCTGTTGAGA 540
DB 1891 AAGCAGTTAGCAGGTTGTGAGCAACCTCCCAACAGTACTTGGGTTTCTGTTGAGA 1950
QY 541 GGTGAGCTCAGACAGGACTAGCTGATTTCTAGGCTGACTAAGAATCCCNAGACCT 600
DB 1951 TGGGGGACTCAGACAGGACTAGCTGATTTCTAGGCTGACTAAGAATCCCNAGACCT 2010
QY 601 ANCTGGGAAGGTGACCGCATCTCTTAAACATGGGGCTTGCACACTTAGCTCACACCCG 660
DB 2011 AGCTGGGAAGGTGACCGCATCTCTTAAACATGGGGCTTGCACACTTAGCTCACACCTG 2070
QY 661 ACCAATCAGAGCTGACTAAATGCTTAATCAGGCACCAAAACAGGAGTAAAGCAATAGCC 720
```

```
2071 ACCAATCAGAGCTGACTAAATGCTTAATAGGCAAAACAGGAGTAAAGAAATAGCC 2130
721 AATCATCTATTGCTGAGACGACAGCGGAAAGCAAGGATTGGGATATAAATCTCAGGCA 780
2131 AATCATCTATTGCTGAGACGACAGCGAGGAGCAATGATCGGATATAAAGCAAGTC 2190
781 TTCAAGCGACGACAGCAACCCCTTTGGTCCCTCCCTCCCTTTGATATGGGAGCTCTGTTT 840
2191 TTGAGCGCGCAACCGCAACCCCTTTGGTCCCTCCCTCCCTTTGATATGGGAGCTCTGTTT 2250
841 CACTCTATTATTCACTCTATTATAATCATGCAACTGCACTCTTCTGGTCCGTGTTTTTATGG 900
2251 CATGCTATTATTCACTCTATTATAATCTTGAACCTGCACTCTTCTGGTCCATGTTTCTACGG 2310
901 CTCAACTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGGCACCGTCAAGACCCGCT 960
2311 CTTGAGCTGAGCTTTGCTCGCCATCCACCACTGCTGTTTGGCGCACCGCAGACCCGCC 2370
961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCTGATCCAGCAGGT 1020
2371 GCTGACTTCCATCCCTTTGGATCATGACAGGTGTCCGCTGTGCTCTGATCCAGCAGGT 2430
1021 ACCATTGCACTCCCGATCAGGCTAAAGGCTTTGCCATTGTTCTCATGGCTTAAGTGCC 1080
2431 ACCATTGCGCTCCCAATTTGGCTTAAGGCTTGCCATTGTTCTCATGGCTTAAGTGCC 2490
1081 TGGGTTTGTCTTAATAAGAACTGAACACTGCTGCTCACTGGGTTTCCATGTTCTCTTCCATGAC 1140
2491 TGGGTTTCACTCTTAATTTGAGTGAACACTAGTCACTGGGTTTCCATGTTCTCTTCTGTGAC 2550
1141 CACGCGCTTCTAATAGAGCTATAA CACTCACGCGATGGCCCAAGATTCATTCCTTTGGTA 1200
2551 CCACAGCTTCTAATAGAGCTATAA CACTCACGCGATGGCCCAAGGTTCCATTCCTTT-GAA 2609
1201 TCTGTGAGCGCAAGAACCCCGCTCAGAGAAAGTGGGCTTGCCACCATTTTGGGAAG 1257
2610 TCCATTAAGCCCAAGAACCCCGCTCAGAGAACACAGAGGTTTGCCACCATTTTGGGAG 2666
```

AF52051982 2694 bp DNA linear PRI 11-FEB-2004
Homo sapiens individual 55 allele B, envelope glycoprotein gene,
complete cds, and 3' long terminal repeat, complete sequence.
AF520520
AF520520.1 GI:33411008

2 of 2
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2694)
Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
The endogenous retroviral locus ERVWE1 is a bona fide gene involved
in hominoid placental physiology
Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
14757825
2 (bases 1 to 2694)
Mallet,F., Bouton,O. and Oriol,G.
Direct Submission
Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée
d'Italie, Lyon 69364 cedex 07, France
Location/Qualifiers
1..2694
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="individual 55 allele B"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-q22"

RESULT 22
AF52051982
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SEGMENT
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source


```
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-q22"
/sex="male"
/cell_type="BMC's"
/notes="Caucasian
endogenous_virus: HERV-W"
misc_feature
43..57
/notes="splice acceptor site"
CDS
284..1900
/notes="syncytin"
/product="envelope glycoprotein"
/protein_id="AAQ17583.1"
/db_xref="GI:33411013"
/translators="MALPYHI,ELFTVL,LPSTLT,APPORCM,TSPPYOFELWRMORP
GNIDAPSYRLSGTPTTAHTHPRNCHYSATLCHNATHYWTGRKNINSPCGGLGV
TVCTYFTQMGSDGGVQDAREKHVKSISOLTRVHGTSSPYKGLDLSKHETLRT
HTRLSVFNLTTLGLHVSANPTNCWICLPLNFRPVYSIPVPEQWNPSTEINTTSV
LVGPLVSNLEITHTSLNLTQKFNSTNTTNSQCIRWTPPTQVCLPSGIFVCGTSA
YRCLNGSSESMCLFPLVPWTIVTEODLYNVIKPRNKRVPILPFIAGVLGALG
TGIGLITTSQFYKLSOELNGDMERVADSLVTLQDLSLAIVLQNRALDLTAE
RGTCFLPGECECYVNSGIVTEKVEIKRDIQRABELRNTGPGLLSQWMPWILP
FLGPLAIIILLFLGFCIFNLLVFSRSIEAVKLQMEPKQSKTKIYRRLDRPASP
RSDVNDIKGTTPPEISAAQPLLRPNAGSS"
misc_feature
1705..1720
/notes="splice acceptor site"
LTR
1959..2694
ORIGIN
Query Match 77.9%; Score 1035.2; DB 9; Length 2694;
Best Local Similarity 90.1%; Pred. No. 3.1e-304;
Matches 1133; Conservative 0; Mismatches 111; Indels 13; Gaps 2;
Qy 1 TCAAAATCGAAGAGCTTTAGACTGTCTAACGGCCAAAGAGGGGGAACCTGTTATTTT 60
Db 1423 TCAAAATCGAAGAGCTTTAGACTGTCTAACGGCTGAAGAGGGGGAACCTGTTATTTT 1482
Qy 61 AGGGAAGAGTGTCTAGTATGTTAATCAATCTGGAATCATTAATGAGAAAGTTAAAGA 120
Db 1483 AGGGAAGAGTGTCTAGTATGTTAATCAATCTGGAATCGTCACTGAGAAAGTTAAAGA 1542
Qy 121 AATTGAGATCGAATAATAATGTAGAGCAGAGGACCTTCAAAACA CTGACCCCTGGGCGCT 180
Db 1543 AATTGAGATCGAATAACA CTGAGCAGAGGAGCTTCGAAACAC TGTGACCCCTGGGCGCT 1602
Qy 181 CCTCAGCAATGAGTGCCTTGGACTCTCCCTCTTTAGGACCTCTAGCAGCTAATAATT 240
Db 1603 CCTCAGCAATGAGTGCCTTGGACTCTCCCTCTTTAGGACCTCTAGCAGCTAATAATT 1662
Qy 241 TTTACTCTCTTTGGACCTGTATCTTCAACTCTCTTTGTTAAGTTTGTCTCTTCCAGAA 300
Db 1663 GCTACTCTCTTTGGACCTGTATCTTAACTCTCTTGTGTTAACTTGTCTCTTCCAGAA 1722
Qy 301 TGAAGCTGTAAAGCTACAATAAGTTCTTCAAAATGGAACCCAGATGCAGTCCATGACTAA 360
Db 1723 CGAAGCTGTAAAGCTA -----CAAATGGAGCCCAAGATGCAGTCCAAAGACTAA 1770
Qy 361 AATCTACCGTGGACCCCTGACCGGCTGTCTAGACTATGCTGTGATGTTAATGACATTGA 420
Db 1771 GATCTACCGAGACCCCTGACCGGCTGTCTAGCCACGATCTGATGTTAATGACATCAA 1830
Qy 421 AGTCACCCCTCCGAGGAAATCTCAACTGCACAAACCCCTACTACCTCCAAATTCAGTAGG 480
Db 1831 AGGACCCCTCTGAGGAAATCTCAGCTGCACAAACCTCTACTAGCCCAATTCAGCAGG 1890
Qy 481 AAGCAGTTAGAGCAGTTGTGAGCCAACTCCCAACAGTACTTGGGTTTCTGTTGAGA 540
Db 1891 AAGCAGTTAGAGCGGTGCTCGGCCAACTCCCAACAGCAGTCTAGGTTTCTGTTGAGA 1950
Qy 541 GGGTGGACTGAGACAGGACTAGCTGGATTCTCCTAGCTGACTAAGAAATCCCAAGCCT 600
Db 1951 TGGGGGACTGAGACAGGACTAGCTGGATTCTCCTAGCTGACTAAGAAATCCCAAGCCT 2010
```

source	1. .2694 /organism="Homo sapiens" /mol_type="genomic DNA" /isolate="individual 65 allele A" /db_xref="taxon:9606" /chromosome="7" /map="7q21-q22" /sex="female" /cell_type="PBMC's" /country="Brazil" /note="Metis endogenous_virus: HERV-W" misc_feature 43..57 /note="splice acceptor site" 284..1900 /note="syncytin" /codon_start=1 /product="envelope glycoprotein" /protein_id="AA017585.1" /db_xref="GI:33411021" /translation="MALPFIHIFLFTVLLPSFTLAPPRCRMTSSPYQEFLLWRMRP GNIDAPSYRLSGTFTFAHTHPRNCHSAATLCMHANTHTYGTGMNIPSCPGILV TVCWTFTOTGMDGGVQDOAREKHVEVISQLTRVHGTSYKGLDLSKLHEITRL HTLVLSENLTLGLHEVSAONPNCWICLPLNFRPVSI PVPQNNFSTENITTSV LVGPLVSNLEITHSNLTCVKFSNTYTTNSQCIRWPTPTQIVCLPSGIFPVCGISA YRLCNGSSSMFLSLFLVPMPTIYTDQDLYSVISPRNKRVPILFPVIGAVIGALG TGIGGTTSTQFYKLSQELNGMDERVADSLVTLQPLNSLAALVILNRRALDILFAE RGCTCLFLAGECCYVYNSGIVTEKVEIKRDRIRRAEELRNTGPMGLLSQWMPILP FLGPLAAILILLGLFCIFNLVNVSSRIEAVKLQWEPKMQSKTKIYRPLDRPAS RSDVNDIKTTPPEISNAQPLLRPSAGSS" 1705..1720 /note="splice acceptor site" 1959..2694
misc_feature	
CDS	
LTR	
ORIGIN	
Query Match	77.98; Score 1035.2; DB 9; Length 2694;
Best Local Similarity	90.18; Pred. No. 3.1e-304;
Matches 1133; Conservative	0; Mismatches 111; Indels 13; Gaps 2;
QY	1 TCAAAATCGAAGAGCTTACAGCTTGTACCGCCMAAGAGGGGGAACCTGTTTATTTT 60
DB	1423 TCAAAATCGAAGAGCTTACAGCTTGTACCGCTGAAGAGGGGGAACCTGTTTATTTT 1482
QY	61 AGGGGAAGATGCTGTATGTTATCAATTCGGAATCATTAATCGAAGAAAGTTAAGA 120
DB	1483 AGGGGAAGATGCTGTATGTTATCAATTCGGAATCATTAATCGAAGAAAGTTAAGA 1542
QY	121 AATTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCCCTGGGCCT 180
DB	1543 AATTGAGATCGAATATAACAGTGTAGAGCAGAGGACCTTCGAAACACTGCACCCCTGGGCCT 1602
QY	181 CCTCAGCAATGGATGCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATT 240
DB	1603 CCTCAGCAATGGATGCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATT 1662
QY	241 TTACTCTCTTTGGACCCCTGTATCTTCAACTTCTTTGTTAAAGTTGTCTCTCCAGAA 300
DB	1663 GCTACTCTCTTTGGACCCCTGTATCTTAACTTCTTTGTTAAAGTTGTCTCTCCAGAA 1722
QY	301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAAACCCAGATGAGTCCATGACTAA 360
DB	1723 CGAAGCTGTAAAGCTA-----CAAAATGGAGCCCAAGATGAGTCCCAAGACTAA 1770
QY	361 AATTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTGTATGTTAATGACATTGA 420
DB	1771 GATCTACCGCAGACCCCTGGACCGGCTGTAGCCACGATCTGATGTTAATGACATCAA 1830
QY	421 AGTCACCCCTCCGAGGAAATCTCAATGTCACAAACCCCTACTACACTCCCAATTCAGTAGG 480
DB	1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACTCTACTAGCCGCCAATTCAGCAGG 1890
QY	481 AAGCAGTTAGAGCAGTTGTGACGCCAACCTCCCAACAGTACTTGGGTTTTCTGTTCAGA 540
DB	

Db	1891 AAGCAGTTAGAGCGGTGCTGCGCAACCTCCCAACAGCAGCACTTAGGTTTTCTGTTCGAGA 1950
QY	541 GGGTGAAGTGAAGAGCAGCACTAGCTGGATTTCTTAGGCTGACTAGAAATCCCAAGCCT 600
DB	1951 TGGGGAGCTGAGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAGAAATCCCAAGCCT 2010
QY	601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATATGGGGCTTGCAACTTAGCTCACAACCCG 660
DB	2011 AGCTGGGAAGGTGACCAATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACAACCTG 2070
QY	661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCGAAAAACAGGAGGTAAAGCAATAGCC 720
DB	2071 ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAGACAGAGGATAAAGAAATAGCC 2130
QY	721 AATCATCTATTGCTGAGAGCAGCGGGAAGAGCAAGATTGGGATATAAATCTCAGGCA 780
DB	2131 AATCATCTATTGCTGAGAGCAGCAGGAGGAGCAATGATCGGATATAAATCTCAGGATC 2190
QY	781 TTCAAGCCAGCAACAGCAACCCCTTTGGTCCCTCCCATTTGATATGGGAGCTCTGTTTT 840
DB	2191 TTGAGCCCGCAACGGCAACCCCTTTGGTCCCTCCCTTTGATATGGGAGCTCTGTTTT 2250
QY	841 CACTCTATTCTCACTCTATTAAATCATCAACTGCACTCTTCTGGTCCGTGTTTTATGG 900
DB	2251 CATGCTATTCTCACTCTATTAAATCTTTGCAACTGCACTCTTCTGGTCCATGTTCTTAGCG 2310
QY	901 CTCAGCTGAGCTTTTGTTCGCATCCACCACTGCTGTTTGGCACCGTCAACAGACCCGCT 960
DB	2311 CTTGAGCTGAGCTTTTCGCTCGCATCCACCACTGCTGTTTGGCCGACCGCAGACCCGCC 2370
QY	961 GCTGACTTCCATCCCTTTTGGATCCAGCAGAGTGTCCACTGTGCTCTGATCCAGCGAGGT 1020
DB	2371 GCTGACTTCCATCCCTTCTGGATCATCGAGGTGTCCGCTGTCTCTGATCCAGCGAGGC 2430
QY	1021 ACCATTGCACTCCCGATCAGGCTAAAGCTTTGCCATTTGCTTCTGATGGTAAAGTGCC 1080
DB	2431 ACCATTGCGCTCCCAATCGGCTAAAGCTTGCCATTTGCTTCTGATGGTAAAGTGCC 2490
QY	1081 TGGTTTGTCTTAATAGAACTGAACACTGCTGCTCACTGGTTCATGTTCTTCCATGAC 1140
DB	2491 TGGTTTCACTCTAATTAAGCTGAACACTAGTCACTGGTTCATGTTCTTCTTCTGTGAC 2550
QY	1141 CCAGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCCTTGTA 1200
DB	2551 CCAGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGTTCCATTCCTT-GAA 2609
QY	1201 TCTGTGAGGCCAAGAACCCAGGTGAGAGGTGAGGCTTGGCACCATTGGGAAG 1257
DB	2610 TCCATAGGCCAAGAACCCAGGTGAGAGGTGAGGAGTGGCACCATTGGGAG 2666

RESULT 25
AF520533S2
LOCUS
DEFINITION
Homo sapiens individual 80 allele A, envelope glycoprotein gene,
complete cds, and 3' long terminal repeat, complete sequence.
AF520534
VERSION
AF520534.1 GI:33411036
KEYWORDS
2 of 2
Homo sapiens (human)
Homo sapiens
Homo sapiens
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS

AF520533S2
2694 bp
DNA
linear
PRI 11-FEB-2004
Homo sapiens individual 80 allele A, envelope glycoprotein gene,
complete cds, and 3' long terminal repeat, complete sequence.
AF520534
GI:33411036
2 of 2
Homo sapiens (human)
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2694)
Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G.,
Bonnaud, B., Lucotte, G., Duret, L. and Mandrand, B.
The endogenous retroviral locus ERVWE1 is a bona fide gene involved
in hominoid placental physiology
Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
14757826
2 (bases 1 to 2694)
Mallet, F., Bouton, O. and Oriol, G.

TITLE Direct Submission
JOURNAL Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
CNRS-bioMérieux, Ecole Normale Supérieure de Lyon - 46 allée
d'Italie, Lyon 69364 cedex 07, France

FEATURES
Location/Qualifiers
source 1..2694
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="individual 80 allele A"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-q22"
/sex="male"
/cell_type="PBMC's"
/note="Caucasian
endogenous_virus: HERV-W"

misc_feature 43..57
/note="splice acceptor site"

CDS
284..1900
/note="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AA017589.1"
/db_xref="GI:33411037"
/translation="MALPYHIFLFTVLLPSFTLTAPPCKMTSSPYQEFLLWRMORP
GNIDAPSRSLSGTPTFTTAHMPKNCYHSATLCMHANTHYWTKMINPSCPGGLGV
TVCWTFQYTGMSGGGQVQAREKHVKEVISQITRVHGTSSPYKGLDKLHETLRT
HTRLVSLFNTLTGLHVSQNPNCWICLPLNRPVSIIPVPEOWNFSTENITTSV
LYGLPLVSLIETHSLNLTVCVFNSTYITNSQCIRWVPTQIQLPSGLFFVCGTSA
LVGNGSESMCFVLPWPTIYTODLSYIVISPKRKRVPILFVIGVIGALG
TGIGTITTFQYKLSQELNGDMRVADSLVTLQDLSLAALVILQNRALLDLTAE
RGCTFLPGECCYQNSQISQVTEKVEIKRDIQRRAELRNTGPMGLLSQMPWTLIP
FLGLPAALIIILLFLPGCIIENLVNFVSSRIEAVKLQWPKWQSKTKIYRRLDRPASP
RSDVNDIKGTPPEISAAQPLLRPNAGSS"

misc_feature 1705..1720
/note="splice acceptor site"

LTR
1959..2694

ORIGIN
Query Match 77.98; Score 1035.2; DB 9; Length 2694;
Best Local Similarity 90.18; Pred. No. 3:1e-304;
Matches 1133; Conservative 0; Mismatches 111; Indels 13; Gaps 2;

Qy 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCCAAAGAGGGGAACTGTTTATTTT 60
Db 1423 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCTGAAAGAGGGGAACTGTTTATTTT 1482

Qy 61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTAATGAGAAAGTTAAAGA 120
Db 1483 AGGGGAAGAATGCTGTTATGTTATGTTAATCAATCGGAATCGTCACTGAGAAAGTTAAAGA 1542

Qy 121 AATTGAGATCGAATAAATGTTAGACAGAGGACCTTCAAAACACTGCACCCCTGGGCGCT 180
Db 1543 AATTGAGATCGAATAAATGTTAGACAGAGGAGCTTCGAAACACTGCACCCCTGGGCGCT 1602

Qy 181 CCTCAGCAATGGATGCCCTGGACTCTCCCTCTTCTAGGACCTCTAGCAGCTATAATATT 240
Db 1603 CCTCAGCAATGGATGCCCTGGACTCTCCCTCTTCTAGGACCTCTAGCAGCTATAATATT 1662

Qy 241 TTACTCCTCTTTGGACCCCTGTATCTTCAACTTCCTTGTAAAGTTGTTCTCTTCCGAAT 300
Db 1663 GCTACTCCTCTTTGGACCCCTGTATCTTAACTCCTTGTAACTTGTCTCTTCCGAAT 1722

Qy 301 TGAAGCTGTAAGCTACAATAATAGTTCTTCAATGGACCCAGATGAGTCCATGACTAA 360
Db 1723 CGAAGCTGTAAGCTACAATAATAGTTCTTCAATGGACCCAGATGAGTCCATGACTAA 1770

Qy 361 AATCTACCGTGGACCCCTGGACCGGCTCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
Db 1771 GATCTACCGACAGACCCCTGGACCGGCTCTGCTAGCCACGATCTGATGTTAATGACATCA 1830

Qy 421 AGTCACCCCTCCGAGGAATCTCAATGTCACACACCCCTACTCACTCCAAATTCAGTAGG 480

Db 1831 AGGCACCCCTCTCTGAGGAATCTCAGCTGCAAACTCTACTACGCCCAATTCAGCAGG 1890
Qy 481 AAGCAGTTAGAGCAGTTGTGAGCAACCTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 540
Db 1891 AAGCAGTTAGAGCGGTGCTCGGCCAACCTCCCAACAGCAGTACTTGGTCTCTGTTGAGA 1950
Qy 541 GGTGAGCTGAGAGACAGGACTAGCTTGTGATTTCTAGGCTGACTAAGAAATCCCAAGCCT 600
Db 1951 TGGGGGACTGAGAGACAGGACTAGCTTGTGATTTCTAGGCTGACTAAGAAATCCCAAGCCT 2010
Qy 601 ANCTGGGAAGGTGACCGCATCTTTTAAACATGGGGCTTGCAACTTAGCTCACACCG 660
Db 2011 AGCTGGGAAGGTGACCAATCCACCTTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2070
Qy 661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAACAGGAGGTAAAGCAATAGCC 720
Db 2071 ACCAATCAGAGAGCTCACTAAATGCTAATAGGCAAAAGCAGGAGGTAAAGCAATAGCC 2130
Qy 721 AATCATCTATTGCTGAGAGACAGCGGGAAGCAAGGATTTGGGATATAAACTCAGGCA 780
Db 2131 AATCATCTATTGCTGAGAGACAGCAGGAGGAGCAATGATCGGATATAAAACCAAGTC 2190
Qy 781 TTCAAGCCGACACAGCAACCCCTTTGGTCCCTCCCTCCCACTTGTATGGGAGCTCTGTTT 840
Db 2191 TTCAAGCCGACCAACGGCAACCCCTTTGGTCCCTCCCTTTGTATGGGAGCTCTGTTT 2250
Qy 841 CACTCTATTCTACTTATTAATCATGCAACTGCACTCTTCTGGTCCGTGTTTATGG 900
Db 2251 CATGCTATTCTACTTATTAATCTTGAACATGCACTCTTCTGGTCCATGTTTCTTAAG 2310
Qy 901 CTCAGCTGAGCTTTTGTGCGCATCCACACATGCTGTTTGGCCACCGTCAAGACCGCT 960
Db 2311 CTTGAGCTGAGCTTTGCTCGCATCCACACATGCTGTTTGGCCACCGCAGACCGCC 2370
Qy 961 GCTGACTTCATCCCTTTGGATCCAGAGAGTTCACATGCTGCTCTGATCCAGCGAGT 1020
Db 2371 GCTGACTTCATCCCTCTCTGATCATGCGGGTGTCCGCTGTCTCTGATCCAGCGAGC 2430
Qy 1021 ACCAATGCACTCCCGATCAGGCTAAAGGCTTGCATTTGCTCTCATGGCTTAAGTGCC 1080
Db 2431 ACCAATGCGCTCCCAATCGGGCTAAAGCTTGCATTTGCTCTCATGGCTTAAGTGCC 2490
Qy 1081 TGGGTTGCTCTAATAGAACTGAACACTGCTGCTCACTGGTTCATGTTCTCTTCCATGAC 1140
Db 2491 TGGGTTTCATCTTAATGAGCTGAACACTAGTCACTGGTTCATGTTCTCTTCTGTGAC 2550
Qy 1141 CCAGGCTCTTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTTCCATTCCTTGGTA 1200
Db 2551 CCAGCTCTTAATAGAGCTATAACACTCACCGCATGGCCCAAGGTTCCATTCCTT-GAA 2609
Qy 1201 TCTGTGAGGCCAAGAACCCCGATCAGAAAGTGTAGGCTTGCCACCATTTGGGAAG 1257
Db 2610 TCCATAAGGCCAAGAACCCCGATCAGAAACAGAGGCTTGCCACCATTTGGGAG 2666

RESULT 26
AF520535S2
LOCUS 2694 bp DNA linear PRI 11-FEB-2004
DEFINITION Homo sapiens individual 80 allele B, envelope glycoprotein gene,
complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION AF520536
VERSION AF520536.1 GI:33411040
KEYWORDS 2 of 2
SEGMENT Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2694)
AUTHORS Mallet F., Bouton O., Prudhomme S., Cheynet V., Oriol G.,
Bonnaud B., Lucotte G., Duret L. and Mandrand B.
TITLE The endogenous retroviral locus ERVWE1 is a bona fide gene involved
in hominoid placental physiology

JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
PUBMED	14757826
REFERENCE	2 (bases 1 to 2694)
AUTHORS	Mallet, F., Bouton, O. and Oriol, G.
TITLE	Direct Submission
JOURNAL	Submitted (07-JUN-2002) Retrovirology Department, UMR 2142 CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée d'Italie, Lyon 69364 cedex 07, France
FEATURES	Location/Qualifiers
source	1..2694
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/isolate="individual 80 allele B"
	/db_xref="taxon:9606"
	/map="7q21-q22"
	/sex="male"
	/cell_type="PBMC's"
	/notes="Caucasian"
	/notes="endogenous_virus: HERV-W"
misc_feature	43..57
	/notes="splice acceptor site"
CDS	284..1900
	/note="syncytin"
	/codon_start=1
	/product="envelope glycoprotein"
	/protein_id="AA017590.1"
	/db_xref="GI:33411041"
	/translation="MALPVIPLFTVLLPSFTLTAPPCKRMWTSPPYOELRMWRP GNIDAPSYSLKGTFTTAAHMPNCHVSHATLKHANTHWYTKMINPSCPGGLV TWCVTFTQMSDGGVQDQARHKVKEVLSQIRKVGTSPPFYKGLDLSKHELTIR HTRLVSLFTTLTLEHVAQNPTNCWICLPLNFRPVYSIPVPEQNNFSTINTSV LVGPLVSNLEITHTLSLTCVKFSNTYTNSQICRWVPTQIVCLPSGIFVFCGTS YRLGSSSMCFPLVPWTIYTBQDLYSVISKPRNRPVILPFIIVAGVIGALG TGIGTITQPFYKLSQELNGDMERVADSLTLODLSLAUVLQNERALDLTAE RGCTFLGEECYVYQSGIVTEKVEIRDRIOREELRNTGPGWLLSQMWPILP FUGPLAAIILLFLFGFCIFNLNLFVSSRIEAVLKQMEPKMQSKTIYRRLDRPASP RSDVNDIKGTPPEEISAAQPLLRPNAGSS"
misc_feature	1705..1720
	/note="splice acceptor site"
LTR	1959..2694
ORIGIN	
Query Match	77.9%; Score 1035.2; DB 9; Length 2694;
Best Local Similarity	90.1%; Pred. No. 3.1e-304;
Matches 1133; Conservative	0; Mismatches 111; Indels 13; Gaps 2;
QY	1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAGAGGGGGAAACCTGTTATTTT 60
DB	1423 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAAACCTGTTATTTT 1482
QY	61 AGGGGAAGATGCTGTATGATCTTAATCAATCTGGAATCATTTACTGAGAAAGTTAAGA 120
DB	1483 AGGGGAAGATGCTGTATGATCTTAATCAATCTGGAATCATTTACTGAGAAAGTTAAGA 1542
QY	121 AATTTCAGATCGAATAATAATGTAGACGAGGAGCTTCAAAAACACTGCACCTCGGGCCCT 180
DB	1543 AATTTCAGATCGAATAATAATGTAGACGAGGAGCTTCAAAAACACTGCACCTCGGGCCCT 1602
QY	181 CCTCAGCAATGGATGCCCTGGACTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATATT 240
DB	1603 CCTCAGCAATGGATGCCCTGGACTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATATT 1662
QY	241 TTTACTCCTCTTTGGACCTGTATCTCACTTCCTTGTAACTTTGTCTCTTCAGAAAT 300
DB	1663 GCTACTCCTCTTTGGACCTGTATCTTAACTCTCTTGTAACTTTGTCTCTTCAGAAAT 1722
QY	301 TGAAGCTGTAAGCTACAAATAGTTCTTTCAAATGGAACCCAGATGCAGTCCATGACTAA 360
DB	1723 CGAAGCTGTAAACTA-----CAATGGAGCCCAAGATGCAGTCCAGACTAA 1770
QY	361 AATTCTACCGTGGACCCCTGACCGGCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420

Db	1771	GATCTACCGACAGCCCTGGACCGGCTGTAGCCACGATCTGATGTTAATGACATCAA	1830
QY	421	AGTCACCCCTCCGAGGAAATCTCAACTGCACAACCCCTACTACTCACTCAATTCAGTAGG	480
Db	1831	AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCAATTCAGCAGG	1890
QY	481	RAGCAGTTAGACAGCTGTGCAGCCAACTCCCCCAACAGTACTTGGTTTCTCTGTGAGA	540
Db	1891	AAGCAGTTAGACAGCGTGTGCGCCAACTCCCCCAACAGCATTAGTTTCTCTGTGAGA	1950
QY	541	GGTGGACTGAGACAGCAGGACTAGCTGGATTTCCTAGGCTGACTAAGAAATCCCAAGCCT	600
Db	1951	TGGGGGACTGAGACAGCAGGACTAGCTGGATTTCCTAGGCTGACTAAGAAATCCCAAGCCT	2010
QY	601	ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGTTCGAACTTAGCTCACAACCG	660
Db	2011	AGCTGGGAAGGTGACCATCCCTTTAAACATGGGGTTCGAACTTAGCTCACAACCTG	2070
QY	661	ACCAATCAGAGAGCTCACTTAAATGCTAATCAGGCAAAACAGGAGGTAAGAAATAGCC	720
Db	2071	ACCAATCAGAGAGCTCACTTAAATGCTAATCAGGCAAAACAGGAGGTAAGAAATAGCC	2130
QY	721	AATCATCTATTGCTGAGACAGCGGGAAGCAAGGATTGGGATATAAACTCAGGCA	780
Db	2131	AATCATCTATTGCTGAGACAGCGGGAAGCAAGGATTGGGATATAAACTCAGGCA	2190
QY	781	TTCAAGCCAGCAACAGCAACCCCTTTGGTCCCTCCCTCCATTTGATGGAGCTCTGTTTT	840
Db	2191	TTCAAGCCAGCAACAGCAACCCCTTTGGTCCCTCCCTCCCTTTGATGGAGCTCTGTTTT	2250
QY	841	CACCTATTCTACTTATTAAATCATGCAACTGCATCTTCTGGTCCGTTTATGCG	900
Db	2251	CATGCTATTCTACTTATTAAATCTTGCACCTGCATCTTCTGGTCCATGTTTCTTACCG	2310
QY	901	CTCAAGCTGAGCTTTTGTTCGCACTCCACCTGCTGTTTGCACCGTCAACAGCCCGCT	960
Db	2311	CTTGGAGCTGAGCTTTGCTGCGCACTCCACCTGCTGTTTGGCCACCGAGACCCACC	2370
QY	961	GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCTGATCAGCGAGGT	1020
Db	2371	GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCGCTGTGCTCTGATCAGCGAGGC	2430
QY	1021	ACCAATGCACTCCGATCAGGCTTAAGGCTTGCCATTTGCTGCTGATGCTTAAGTGCC	1080
Db	2431	ACCAATGCGCTCCCAATCGGCTTAAGGCTTGCCATTTGCTGCTGATGCTTAAGTGCC	2490
QY	1081	TGGGTTTGTCTTAATAGAACTCAACACTGCTCACTGGTTCATGTTCTCTTCCATGAC	1140
Db	2491	TGGGTTTGTCTTAATAGAACTCAACACTGCTCACTGGTTCATGTTCTCTTCTGTGAC	2550
QY	1141	CCACGCTTCTTAATAGAGCTATAAACAACCTCACCGCATGGCCCAAGATTCATCTCTTGGTA	1200
Db	2551	CCACGCTTCTTAATAGAGCTATAAACAACCTCACCGCATGGCCCAAGATTCATCTCTT	2609
QY	1201	TTGTGAGCCCAAGAACCCAGGTGAGAAAGTGGCTTGGCCACCATTTGGGAAG	1257
Db	2610	TCATAAGCCCAAGAACCCAGGTGAGAAACACGAGGCTTGCCACCATTTGGGAG	2666

RESULT 27	
AF520537S2	2694 bp DNA linear PRI 11-FEB-2004
LOCUS	Homo sapiens individual 81 allele A, envelope glycoprotein gene,
DEFINITION	complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION	AF520538
VERSION	AF520538.1 GI:33411044
KEYWORDS	2 of 2
SEGMENT	Homo sapiens (human)
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
	1 (bases 1 to 2694)

AUTHORS Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
 Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
 TITLE The endogenous retroviral locus ERVWE1 is a bona fide gene involved
 in hominoid placental physiology
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
 PUBMED 14757826
 REFERENCE 2 (bases 1 to 2694)
 AUTHORS Mallet,F., Bouton,O. and Oriol,G.
 TITLE Direct Submission
 JOURNAL Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
 CNRS-bioMérieux, Ecole Normale Supérieure de Lyon - 46 allée
 d'Italie, Lyon 69364 cedex 07, France

FEATURES
 source
 1..2694
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /isolate="individual 81 allele A"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7q21-q22"
 /sex="female"
 /cell_type="PBMC's"
 /notes="endogenous_virus: HERV-W"
 misc_feature
 43..57
 /notes="splice acceptor site"
 CDS
 284..1900
 /notes="syncytin"
 /codon_start=1
 /product="envelope glycoprotein"
 /protein_id="AAQ17591.1"
 /db_xref="GI:33411045"
 /translations="MALPHIFLFTLLPSFTLAPPKRCMTSSPYQEFLLWRMORP
 TVCWYFTOTGMSDGGVDOAREKHVKEVISOLTRVHGTSPPYKGLDLSKLHETLRT
 GNIDAPSYRLSGTPTFAHTHMPRNCHVSHATLCHNANTHYMTGKMINPSCPGGLGV
 HTPLVSLFNTLLGLHEVSAQNTNCWICLPLNRPYVSIPIVPEQWNNSTEINTSV
 LVGLNLSSEMTISNLTQVKNFTSYITNSQCIRWPTPTQIVCLPSPGFFVCGTSA
 YRLCNGSSSEFLSLVLPYITQDLYSVISKPRNKRVPILPFVIGAGVIGALG
 RGIGTITTFQFYKLSQELNGMVDVSLVTLQDLNSLAALVILNRRALDLLTAE
 TGGCTFLGEBECYVYNSGIVTEKVEIKRDRIRRAEELRNTGPMGLLSQWMPWTLF
 FLGLAAIILLFLFGPICFNLNVFSSRIEAVKLQWEPKMQSKTKIYRPLDRPASP
 RSDVDIKGTPPEISNAQPLLRPSAGSS"
 misc_feature
 1705..1720
 /notes="splice acceptor site"
 LTR
 1959..2694

ORIGIN
 Query Match 77.98; Score 1035.2; DB 9; Length 2694;
 Best Local Similarity 90.18; Pred. No. 3.1e-304;
 Matches 1133; Conservative 0; Mismatches 111; Indels 13; Gaps 2;
 1 TCAAAATCGAAGAGCTTGTAGACTTGCTAACCGCCAAAGAGGGGAACTGTTTATTTT 60
 1423 TCAAAATCGAAGAGCTTGTAGACTTGCTAACCGCTGAAGAGGGGAACTGTTTATTTT 1482
 61 AGGGGAAGATGCTGTTATGTTATGTTATCAATTCGAATCATTTACTGAGAAAGTTAAAGA 120
 1483 AGGGGAAGATGCTGTTATGTTATGTTATCAATTCGAATCGGAAATCGTCACTGAGAAAGTTAAAGA 1542
 121 AATTGAGATCGAATATGTTAGACGAGAGGACCTTCAAAACACTGCACCCCTGGGCCT 180
 1543 AATTGAGATCGAATATGTTAGACGAGAGGAGCTTCGAAACACTGGACCCCTGGGCCT 1602
 181 CCTCAGCAATGGATGCCCTGGACTCTCCCTCTTCTAGGACCTCTAGCAGCTATAATATT 240
 1603 CCTCAGCAATGGATGCCCTGGATCTCCCTCTTCTAGGACCTCTAGCAGCTATAATATT 1662
 241 TTACTCTCTTTGGACCCCTGTAFTCTCAACTTCCTTGTAAAGTTGTTCTTCCAGAAT 300
 1663 GCTACTCTCTTTGGACCCCTGTAFTCTCAACTTCCTTGTAAAGTTGTTCTTCCAGAAT 1722
 301 TGAAGCTGTAAGCTACAATAGTTCTTCAATGGACCCCAAGCTCGATCCATGACTAA 360
 1723 CGAAGCTGTAAGCTA-----CAATGGAGCCCAAGATGAGTCCAAGACTAA 1770

QY 361 AATCTACCGTGGACCCCTGGACCGGCTCTGTAGACTATGCTCTGTATGTTAATGACATTGA 420
 DB 1771 GATCTACCGCAGACCCCTGGACCGGCTCTGTAGCCACGATCTGTATGTTAATGACATCA 1830
 QY 421 AGTCACCCCTCCGAGGAAATCTCAATCGACAAACCCCTACTACACTCCAATTCAGTAGG 480
 DB 1831 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAAACCTCTACTACGCCCAATTCAGCAG 1890
 QY 481 AAGCAGTTAGAGCAGTTGTTCAGCAACCTCCCAACAGTACTTGGGTTTTCCTGTTTGA 540
 DB 1891 AAGCAGTTAGAGCGGTGTCGGCCAACTCCCAACAGCACTTAGGTTTTCCTGTTTGA 1950
 QY 541 GGGTGACTGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAAGAATCCCAAGACCT 600
 DB 1951 TGGGGACTGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAAGAATCCCAAGACCT 2010
 QY 601 ANCTGGGAAGTGACCGCATCCATCTTTAAACATGCGGCTTGCACACTTAGCTCACACCG 660
 DB 2011 AGCTGGGAAGTGACCAATCCACCTTTAAACACAGGGGCTTGCAACTTAGCTCACACCTG 2070
 QY 661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAACAGGAGTAAAGCAATAGCC 720
 DB 2071 ACCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAAACAGGAGTAAAGCAATAGCC 2130
 QY 721 AATCATCTATTGCTTGAGACAGCGGGAAGGACAAAGGATTTGGGATATAAACTCAGGCA 780
 DB 2131 AATCATCTATTGCTTGAGACAGCGGAGGACAAATGATCGGATATAAAACCAAGTC 2190
 QY 781 TTCAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGATAGGAGCTCTGTTTT 840
 DB 2191 TTGAGCCGGCAACGGCAACCCCTTTGGGTCCCTCCCTTTGATAGGAGCTCTGTTTT 2250
 QY 841 CACTCTATTCTACTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGCTTTTATG 900
 DB 2251 CATGCTATTCTACTCTATTAAATCTTGAACATGCACTCTTCTGGTCCATGTTCTTACGG 2310
 QY 901 CTCAACTGAGCTTTTGTTCGCATCTCACCACTGCTGTTTGGCACCGTCACAGACCCGCT 960
 DB 2311 CTTGAGCTGAGCTTTGCTCGCATCCACCACTGCTGTTTGGCCACCGCAGACCCGCT 2370
 QY 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCTGATCCAGCGAGT 1020
 DB 2371 GCTGACTTCCATCCCTCTGGATCATGAGGCTGTCGGCTGTGCTCTGATCCAGCGAGC 2430
 QY 1021 ACCATTGGCACTCCCGATCAGGCTAAAGGCTTGCATTTGCTCATGGCTAAAGTGC 1080
 DB 2431 ACCATTGGCTTCCCAATCGGCTAAAGGCTTGCATTTGCTCATGGCTAAAGTGC 2490
 QY 1081 TGGGTTTGTCTTAATAGAACTGAACACTGGTCACTGGGTTCCATGTTCTTCTCCATGAC 1140
 DB 2491 TGGGTTTCACTTAATTTGAGCTGAACACTAGTCACTGGGTTCCATGTTCTTCTCTGAC 2550
 QY 1141 CCAGCGCTCTTAATAGAGCTATAACTCACCAGCATGGCCCAAGATTCCATTCCTTGGTA 1200
 DB 2551 CCAGCTCTTCTTAATAGAGCTATAAACAACCTCACCAGCATGGCCCAAGTTCCATTCCTT-GAA 2609
 QY 1201 TCTGTGAGCCCAAGAACCCCGATCAGGAANGTGAAGGCTTGCCACCATTTGGGAAG 1257
 DB 2610 TCATTAAGCCCAAGAACCCCGATCAGGAACACAGAGGCTTGCCACCATTTGGGAG 2666

RESULT 28
 AF52054982
 LOCUS 2694 bp DNA linear PRI 11-FEB-2004
 DEFINITION Homo sapiens individual 95 allele A, envelope glycoprotein gene,
 complete cds, and 3' long terminal repeat, complete sequence.
 ACCESSION AF520550
 VERSION AF520550.1 GI:33411068
 KEYWORDS
 2 of 2
 SEGMENT Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 2694)

REFERENCE
AUTHORS Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G., Bonnaud, B., Lucotte, G., Duret, L. and Mandrand, B.

TITLE The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)

PUBMED 14757826

REFERENCE 2 (bases 1 to 2694)

AUTHORS Mallet, F., Bouton, O. and Oriol, G.

TITLE Direct Submission

JOURNAL Submitted (07-JUN-2002) Retrovirology Department, UMR 2142 CNRS-Bioherieux, Ecole Normale Supérieure de Lyon - 46 allée d'Italie, Lyon 69364 cedex 07, France

FEATURES
source
1: 2694
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="individual 95 allele A"
/db_xref="taxon:9606"
/map="7q21-q22"
/sex="female"
/cell_type="PBMC's"
/notes="endogenous_virus: HERV-W"
misc_feature
43..57
/notes="splice acceptor site"
CDS
284..1900
/notes="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAQ17597.1"
/db_xref="GI:33411069"
/translation="MALPHIFLFTLLPSFTLTAPPCCMTSSPYOEFELRMORP
GNIDAPSRSLSGTPTFAHTHMPNCHSHATLCHNANTHTYGTGRWNPSCGGLV
TCWTFITQMSDGGVQDAREKHVEVISQLRVHGTSSPYSGIDLSKHEHETUR
HTRLVSLNTEITGLHVSQNPNTNCWICLPNFRPVSVIPVEQWNPSTEINTTSV
LVGLVSNLEITHTSNLTCKFNSTYNTNSQIRWTPPTQIVCLPSGIFVFCGTS
YRLGSSMSCLPSLPMPMTITVQDLYSYVISKPNRNPVILPFPVIGAGVIGALG
TCGGITTSQFYKLSQELNGDMERVADSLVTLQDLNSLAAVLQNRALDLTAE
RGCTCLFLGECYVQSGIVTEKYKEIRDIQRAABELRNTGPGWLLSQMPWILP
FLGLAAIILLGLFPCIFNLVNFVSSRIEFLVQNEPMPQMSKTIYRPLDRPASP
RSDVNDIKGTPPBEISAAQPLLRNSAGS"
misc_feature
1705..1720
/notes="splice acceptor site"
LTR
1959..2694

ORIGIN

Query Match 77.9%; Score 1035.2; DB 9; Length 2694;
Best Local Similarity 90.1%; Pred. No. 3.1e-304;
Matches 1133; Conservative 0; Mismatches 111; Indels 13; Gaps 2;

QY 1 TCAAAATCGAAGAGCTTTAGACTGCTAACCGCCAAAGAGGGGGAACCTGTTATTTT 60
DB 1423 TCAAAATCGAAGAGCTTTAGACTGCTAACCGCTGAAAGAGGGGGAACCTGTTATTTT 1482

QY 61 AGGGAAGAATGCTGTTAGTATGTTAATCAATCGGAATCATTAATCTGAGAAAGTTAAAGA 120
DB 1483 AGGGAAGAATGCTGTTATGTTATGTTAATCAATCGGAATCGTCACTGAGAAAGTTAAAGA 1542

QY 121 AATTTGAGATCGAATAATAATGTAGAGCAGAGGACCTTCAAAACACATGACCCCTGGGCGCT 180
DB 1543 AATTCGAGATCGAATAACAACGTAGAGCAGAGGAGCTTCGAAACACATGACCCCTGGGCGCT 1602

QY 181 CCTCAGCAATGGAATGCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATTT 240
DB 1603 CCTCAGCAATGGAATGCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATTT 1662

QY 241 TTTACTCTCTTTGGACCCGTGATCTTCAACTCTCTGTTAAGTTGCTCTCTCCAGAAT 300
DB 1663 GCTACTCTCTTTGGACCCGTGATCTTCAACTCTCTGTTAAGTTGCTCTCTCCAGAAT 1722

QY 301 TGNAGCTGTAAAGCTTACAAATAGTCTTTCAATAGGAACCCAGATCGAGTCCATGACTAA 360
DB 1723 CGAAGCTGTAAAGCTTACAAATAGTCTTTCAATAGGAACCCAGATCGAGTCCATGACTAA 1770

QY 361 AATCTACCGTGGACCCCTCGACCGGCTGTCTAGACTATGCTCTGATGTTTAATGACATTTGA 420
DB 1771 GATCTACCGCAGACCCCTCGACCGGCTGTCTAGCCACACGATCTGATGTTTAATGACATCA 1830

QY 421 AGTCACCCCTCCGAGGAAATCTCAATGTCACAAACCCCTACTACATCTCCAATTCAGTAGG 480
DB 1831 AGGCACCCCTCCTCGAGGAAATCTCAGCTGCACAAACCTCTACTACGCGCCCAATTCAGCAGG 1890

QY 481 AAGCAGTTAGACGAGTTGTTCAGCCACCTCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
DB 1891 AAGCAGTTAGACGAGTTGTTCAGCCACCTCCCAACAGTACTTGGGTTTTCCTGTTGAGA 1950

QY 541 GGGTGACTGAGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCCAAGACCT 600
DB 1951 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCCAAGACCT 2010

QY 601 ANCTGGAAAGGTGACCGCATCCATTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
DB 2011 AGCTGGAAAGGTGACCGCATCCATTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 2070

QY 661 ACCAATCAGAGAGCTCACTAATAATGCTAATCAGGCAAAAACAGGAGTAAGCAATAGCC 720
DB 2071 ACCAATCAGAGAGCTCACTAATAATGCTAATCAGGCAAAAACAGGAGTAAGCAATAGCC 2130

QY 721 AATCATCTATTGCTGAGAGACAGCGGGAAGGACAGGATTCGGGATATAAATCTCAGGCA 780
DB 2131 AATCATCTATTGCTGAGAGACAGCGGGAAGGACAGGATTCGGGATATAAATCTCAGGATC 2190

QY 781 TTCAAGCCAGCAACAGCAACCCCTTTGGTCCCTCCCTCCCATTTGATGGGAGCTCTGTTTT 840
DB 2191 TTCAAGCCAGCAACAGCAACCCCTTTGGTCCCTCCCTCCCATTTGATGGGAGCTCTGTTTT 2250

QY 841 CACTCTATTCTACTCTAATAATCATGCAATGCACTCTCTGGTCCGTTGTTTATG 900
DB 2251 CATGCTATTCTACTCTAATAATCATGCAATGCACTCTCTGGTCCGTTGTTTATG 2310

QY 901 CTCACGCTGAGCTTTTGTTCGCCATCCACCTGCTGTTTGGCCACCGTACAGACCCGCT 960
DB 2311 CTTGAGCTGAGCTTTTGTTCGCCATCCACCTGCTGTTTGGCCACCGTACAGACCCGCT 2370

QY 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTCTCTGATCCAGCGAGGT 1020
DB 2371 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTCTCTGATCCAGCGAGGT 2430

QY 1021 ACCATTGGCACTCCCGATCAGGCTTAAGGCTTGCATTTGTTCTCTGATGGCTAAGTGC 1080
DB 2431 ACCATTGGCGCTCCCAATCGGCTTAAGGCTTGCATTTGTTCTCTGATGGCTAAGTGC 2490

QY 1081 TGGGTTTGTCTTAATAGAACTGAACACTGGTCACTGGGTTTCCATGTTCTCTTCCATGAC 1140
DB 2491 TGGGTTTGTCTTAATAGAACTGAACACTGGTCACTGGGTTTCCATGTTCTCTTCTGAC 2550

QY 1141 CCACGGCTTCTTAATAGAGCTATAAACAACCTCACCGCATGGCCCAAGATTCCATCTTGTA 1200
DB 2551 CCACAGCTTCTTAATAGAGCTATAAACAACCTCACCGCATGGCCCAAGATTCCATCTT-GAA 2609

QY 1201 TCTGTGAGGCCAAGAACCCAGGTACAGAAAGTGGGCTTGCCACCATTTGGGAAG 1257
DB 2610 TCCATAAGGCCAAGAACCCAGGTACAGAAACAGAGGCTTGCCACCATTTGGGAG 2666

RESULT 29
AF520553S2
LOCUS
DEFINITION Homo sapiens individual 96 allele A, envelope glycoprotein gene, complete cds, and 3' long terminal repeat, complete sequence.
ACCESSION AF520554
VERSION AF520554.1
GI:33411076
KEYWORDS

```

SEGMENT 2 of 2
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2694)
AUTHORS   Mallet,F., Bouton,O., Prudhomme,S., Cheynet,V., Oriol,G.,
          Bonnaud,B., Lucotte,G., Duret,L. and Mandrand,B.
          The endogenous retroviral locus ERVWE1 is a bona fide gene involved
          in hominoid placental physiology
          Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
JOURNAL   14757826
PUBMED
REFERENCE 2 (bases 1 to 2694)
AUTHORS   Mallet,F., Bouton,O. and Oriol,G.
          Direct Submission
          Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
          CNRS-bioMérieux, Ecole Normale Supérieure de Lyon - 46 allée
          d'Italie, Lyon 69364 cedex 07, France
FEATURES
    source
        1..2694
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /isolate="individual 96 allele A"
            /db_xref="taxon:9606"
            /chromosome="7"
            /map="7q21-q22"
            /sex="male"
            /cell_type="PBMC's"
            /note="endogenous_virus: HERV-W"
        43..57
            /note="splice acceptor site"
        284..1900
            /note="syncytin"
            /codon_start=1
            /product="envelope glycoprotein"
            /protein_id="AAQ17599.1"
            /db_xref="GI:33411077"
            /translation="MALPFYHIFLTVLLPSFTLTAPPCKMTSSPYQBEFLWRMORP
            GNIDAPSYRLSGTPTFTAHTRPNRCHVSATLCMHANTHYWTGMINPSCPGILGV
            TVCWYPTQGTMDGGVQDQAREKHVEKVISOLTRVHGTSPPYKGLDLSKLHETLRT
            HTRLVSLNTTGLHEVSAQNTNCWICLPNFRPVVSPVPEQNNFSTEINTSV
            LVGPLVSNLEITHSNLTCVKPSNTYTNNSQICIRWVPTQIYVCLPSGIFVCGTSA
            YRCNGSSKSFSLFVPMYTIYBODLYSVISPRNKRVPFLFVIGVIGAG
            TGGTSTQFYKLSQELNGDMERVADSLVTLQDLSLAVALVLRNRALLDLTAE
            RGCTFLGECYQNSQIVTEKVEIRDIQRRAELRNTGPGWLSQWMPILP
            FLGPLAAIILLLFGPCIFNLVNFVSSRIEAVKLQWEPKMQSKTKIYRRLDRPAS
            RSDVNDIKGTPPEISAAQPLLRPNAGS"
        1705..1720
            /note="splice acceptor site"
        1959..2694
            /note="splice acceptor site"
    CDS
        1..1900
            /note="syncytin"
            /product="envelope glycoprotein"
            /protein_id="AAQ17599.1"
            /db_xref="GI:33411077"
            /translation="MALPFYHIFLTVLLPSFTLTAPPCKMTSSPYQBEFLWRMORP
            GNIDAPSYRLSGTPTFTAHTRPNRCHVSATLCMHANTHYWTGMINPSCPGILGV
            TVCWYPTQGTMDGGVQDQAREKHVEKVISOLTRVHGTSPPYKGLDLSKLHETLRT
            HTRLVSLNTTGLHEVSAQNTNCWICLPNFRPVVSPVPEQNNFSTEINTSV
            LVGPLVSNLEITHSNLTCVKPSNTYTNNSQICIRWVPTQIYVCLPSGIFVCGTSA
            YRCNGSSKSFSLFVPMYTIYBODLYSVISPRNKRVPFLFVIGVIGAG
            TGGTSTQFYKLSQELNGDMERVADSLVTLQDLSLAVALVLRNRALLDLTAE
            RGCTFLGECYQNSQIVTEKVEIRDIQRRAELRNTGPGWLSQWMPILP
            FLGPLAAIILLLFGPCIFNLVNFVSSRIEAVKLQWEPKMQSKTKIYRRLDRPAS
            RSDVNDIKGTPPEISAAQPLLRPNAGS"
        1705..1720
            /note="splice acceptor site"
        1959..2694
            /note="splice acceptor site"
    LTR
        1959..2694
            /note="splice acceptor site"
    ORIGIN
        Query Match 77.98; Score 1035.2; DB 9; Length 2694;
        Best Local Similarity 90.18; Pred. No. 3.le-304;
        Matches 1133; Conservative 0; Mismatches 111; Indels 13; Gaps 2;
        QY 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCAGCGCAAGAGGGGAACTGTTATTTT 60
        DB 1423 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCAGCGCAAGAGGGGAACTGTTATTTT 1482
        QY 61 AGGGGAAGATGCTGTTAGTATGTTAATCAATCTGGAATCATTAATGAGAAAGTTAAGA 120
        DB 1483 AGGGGAAGATGCTGTTATGTTATGTTAATCAATCGGAATCGTCACTGAGAAGTTAAGA 1542
        QY 121 AATTGAGATCGAATAAATGATAGAGCAGAGGACCTTCAAAACACTGACCTGGGGCCT 180
        DB 1543 AATTGAGATCGAATAAATGATAGAGCAGAGGACCTTCAAAACACTGACCTGGGGCCT 1602
        QY 181 CCTCAGCAATGATGCGCTGGAATCTCCCTCTCTTAGGACCTCTTAGCAGCTATAATATT 240
        DB 1603 CCTCAGCAATGATGCGCTGGAATCTCCCTCTCTTAGGACCTCTTAGCAGCTATAATATT 1662
        QY 241 TTACTCTCTTTGGACCCCTGATCTTCAACTTCTTGTGTTAAGTTGCTCTCTCCAGAA 300

```

RESULT 30
AF520561S2
LOCUS
DEFINITION

AP520561S2 2694 bp DNA linear PRI 11-FEB-2004
Homo sapiens individual 71 allele A, envelope glycoprotein gene,
complete cds, and 3' long terminal repeat, complete sequence.

```

1663 GCTACTCCTCTTTGGACCCCTGTATCTTTAACCTCTTGTGTAACTTTGTCTCTTCCAGAA 1722
301 TGAAGCTGTAAAGCTACAAATAGTCTTCTCAAATGGAACCCAGATGCAGTCCATGACTAA 360
1723 CGAAGCTGTAAACTA-----CAATGGAGCCCAAGATGCAGTCCCAAGACTAA 1770
361 AATCTACCGTGGACCCCTGAGACCGGCTGTAGACTATGCTGTGATGTTAATGACATGGA 420
1771 GATCTACCGCAGACCCCTGAGACCGGCTGTAGACCCAGCATCTGATGTTAATGACATCA 1830
421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAAACCCCTACTACATCTCCAAATTCAGTAG 480
1831 AGCACCCCTCTTGGAGAAATCTCAGCTGCACAACTCTACTACGCCCAATTCAGCAGG 1890
481 AAGCAGTTAGACGAGTTGTTCAGCAACCTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 540
1891 AAGCAGTTAGAGGGTGTCTGCGCAACCTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 1950
541 GGCTGAGCTGAGACGAGCTAGCTGGATTTCCTTAGGCTGACTAAGATCCCNAAAGCCT 600
1951 TGGGGGACTGAGACGAGCTAGCTGGATTTCCTTAGGCTGACTAAGATCCCTAAGCCT 2010
601 ANCTGGGAAGTGCACCGCATCCATCTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
2011 AGCTGGGAAGTGCACCATCCATCTTAAACACAGGGGCTTGCAACTTAGCTCACACCTG 2070
661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAACAGGAGGTAAAGCAATAGCC 720
2071 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAACAGGAGGTAAAGCAATAGCC 2130
721 AATCATCTATTGCTGAGACGAGCGGGAAGCAAGGATTCGGGATATAAACTCAGGCA 780
2131 AATCATCTATTGCTGAGACGAGCGGGAAGCAAGGATTCGGGATATAAACTCAGGCA 2190
781 TTCAAGCAGACACAGCAACCCCTTTGGTCCCTCCCACTTGTATGGGAGCTCTGTTTT 840
2191 TTCAAGCAGCAGCAACCCCTTTGGTCCCTCCCTCTTGTATGGGAGCTCTGTTTT 2250
841 CACTCTATTCTACTCTATTAAATCATGCAACTGCACTCTTCTGCTCGTGTGTTTATGG 900
2251 CATGCTATTCTACTCTATTAAATCTTGAACATGCACTCTTCTGCTCGTGTGTTTACGG 2310
901 CTCAAGCTGAGCTTTTGTGCGCATCCACACTGCTGTTTGGCCACCGTACAGACCCGCT 960
2311 CTTGAGCTGAGCTTTTGTGCGCATCCACACTGCTGTTTGGCCACCGTACAGACCCGCT 2370
961 GCTGACTTCCATCTTGTGATCCAGAGGTGTCACCTGCTGCTGATCCAGCGAGGT 1020
2371 GCTGACTTCCATCTTGTGATCCAGAGGTGTCACCTGCTGCTGATCCAGCGAGGT 2430
1021 ACCCATTTGCCACTTCCCGATCAGGCTAAAGGCTTGGCAATTTCTCTGATGGCTAAGTGC 1080
2431 ACCCATTTGCCACTTCCCGATCAGGCTAAAGGCTTGGCAATTTCTCTGATGGCTAAGTGC 2490
1081 TGGGTTGCTCTAATAGACTGAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
2491 TGGGTTGCTCTAATAGACTGAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2550
1141 CCAGGCTTCTAATAGACTGAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
2551 CCAGGCTTCTAATAGACTGAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2609
1201 TCTGTGAGGCGCAAGAACCCCGATCAGAGGCTGAGGCTTGGCCACCATTTGGGAAG 1257
2610 TCCATAGGCGCAAGAACCCCGATCAGAGGCTTGGCCACCATTTGGGAAG 2666

```

```

ACCESSION AF520562
VERSION AF520562.1 GI:33411092
KEYWORDS
SEGMENT
SOURCE
ORGANISM Homo sapiens (human)
2 of 2
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2694)
AUTHORS Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G.,
Bonnaud, B., Lucotte, G., Duret, L. and Mandrand, B.
TITLE The endogenous retroviral locus ERVW1 is a bona fide gene involved
in hominoid placental physiology
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 101 (6), 1731-1736 (2004)
PUBMED 14757826
REFERENCE 2 (bases 1 to 2694)
AUTHORS Mallet, F., Bouton, O. and Oriol, G.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2002) Retrovirology Department, UMR 2142
CNRS-bioMerieux, Ecole Normale Supérieure de Lyon - 46 allée
d'Italie, Lyon 69364 cedex 07, France
FEATURES
source
1..2694
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="individual 71 allele A"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-q22"
/sex="male"
/cell_type="PBMC's"
/note="African"
endogenous_virus: HERV-W
misc_feature 43..57
/note="splice acceptor site"
CDS
284..1900
/note="syncytin"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AAQ17603.1"
/db_xref="GI:33411093"
/translation="MALPYHIFLFTVLLPSFTLTAPPICRMTSSPYQBLRMORP
GNIDAPYSLSGTPFTTAHMRNCHYSATLCHMANTHYWTKMINPSCFGLGV
TVCTWYPTSGDGGVQDQREKHEVISQITRVHGTSSPYKGLDLSKLHETLRT
HTRLVSLNTLTHTSNLTCVKFSNTVTYNSQICRTWVTPQIVCLPSGIFPVCSTSA
LVPLVSNLBIHTTSLNLTQVFNSTVYNSQICRTWVTPQIVCLPSGIFPVCSTSA
YRLNGSSEMCFLSFLVPWTLYTEODLYNVI SKPRNRPVLPVIGAVLGAIG
TGIGITITSTQFYKLSQELNGDMERVADSLVLDQNLNLAIVLQNRALLLTAE
RGSTCLFLGEECCYYNQSIGVTEKVEIEDRIQRRABELRNTGPMGLSOWMPWILP
FLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKVQSKTKIYRPLDRPAS
RSDVNDIKGTPPEISAAQPLLRPNSSAGS"
misc_feature 1705..1720
/note="splice acceptor site"
ORIGIN
1705
1959..2694
Query Match 77.9%; Score 1035.2; DB 9; Length 2694;
Best Local Similarity 90.1%; Pred. No. 3.1e-304;
Matches 1133; Conservative 0; Mismatches 111; Indels 13; Gaps 2;
Qy 1 TCRAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAGAGGGGGAACCTGTTTATTTT 60
Db 1423 TCRAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAAGAGGGGGAACCTGTTTATTTT 1482
Qy 61 AGGGGAAGATGCTGTATGATGTTATCAATCTGGAATCAATTAAGTGAAGTTAAGA 120
Db 1483 AGGGGAAGATGCTGTATGATGTTATCAATCTGGAATCGTCACTGAGAAGTTAAGA 1542
Qy 121 AATTGAGATCGAATAATGCTAGCAGAGGAGCTTCAAACACTGACCCCTGGGCGCT 180
Db 1543 AATTGAGATCGAATAATGCTAGCAGAGGAGCTTCAAACACTGACCCCTGGGCGCT 1602
Qy 181 CCTCAGCAATGGATGCGCTGGAAGCTCTCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240

```

Search completed: January 22, 2005, 16:14:04

```

Db 1603 CCTCAGCAATGGATGCGCTGGATTCCTCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1662
Qy 241 TTTACTCTCTTTGGACCCCTGTATCTTCAACTTCCTTGTAAAGTTTGTCTCTTCCAGAAAT 300
Db 1663 GCTACTCTCTTTGGACCCCTGTATCTTAAACCTCTTGTAAAGTTTGTCTCTTCCAGAAAT 1722
Qy 301 TGAAGCTGTAAAGCTACAATAATAGTTCTTCAAAATGGAAACCCAGATCGAGTCCCATGACTAA 360
Db 1723 CGAAGCTGTAAAGCTACAATAATAGTTCTTCAAAATGGAAACCCAGATCGAGTCCCATGACTAA 1770
Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGTCTAGACTATGCTGTAGTGTAAATGACATGTA 420
Db 1771 GATCTACCGAGACCCCTGGACCGGCTGTCTAGCTCCACAGATCTGATGTAAATGACATCAA 1830
Qy 421 AGTCACCCCTCCGAGGAAATCTCAATGTGCACAAACCCCTACTACACTCTCAATTCAGTAGG 480
Db 1831 AGGCACCCCTCTCTGAGGAAATCTCAGCTGCACAAACCTCTACTACGCCCCAATTCAGCAGG 1890
Qy 481 AAGCAGTTAGAGCAGTTGTTCAGCCAACTCCCAACAGTACTTTGGGTTTCTTCTGTTGAGA 540
Db 1891 AAGCAGTTAGAGCAGTTGTTCAGCCAACTCCCAACAGTACTTTGGGTTTCTTCTGTTGAGA 1950
Qy 541 GGGTGACTGAGAGACGAGTACTAGCTGGATTTCTCTAGGCTGACTAAGAAATCCCNAAAGCCT 600
Db 1951 TGGGGACTGAGAGACGAGTACTAGCTGGATTTCTCTAGGCTGACTAAGAAATCCCNAAAGCCT 2010
Qy 601 ANCTGGGAAGGTGACCGCATCCAATCTTTAAACAATGGGGCTTGCAACTTAGCTCACACCCG 660
Db 2011 AGCTGGGAAGGTGACCGCATCCAATCTTTAAACAATGGGGCTTGCAACTTAGCTCACACCTG 2070
Qy 661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCGAAGAAACAGGAGGTAAGCAATAGCC 720
Db 2071 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCGAAGAAACAGGAGGTAAGCAATAGCC 2130
Qy 721 AATCATCTATTGCTCAGAGACAGCGGGAAGCAAGGATTCGGGATATAAACTCAGGCA 780
Db 2131 AATCATCTATTGCTCAGAGACAGCGGGAAGCAAGGATTCGGGATATAAACTCAGGCA 2190
Qy 781 TTCAAGCCAGCAACAGCAACCCCTTTGGTCCCTCCCTCCCAATTTGATGGGAGCTCTGTTTT 840
Db 2191 TTCAAGCCAGCAACAGCAACCCCTTTGGTCCCTCCCTCCCTTTGATGGGAGCTCTGTTTT 2250
Qy 841 CACTCTATTTCACTCTATTAATCATGCAAACTGCACTCTCTCTGGTCCGTTGTTTATGG 900
Db 2251 CATGCTATTTCACTCTATTAATCATGCAAACTGCACTCTCTCTGGTCCGTTGTTTATGG 2310
Qy 901 CTCAAGCTGAGCTTTTGTGGCCATCCACCACTGCTGTTTGGCCACCGTCAAGACCCGCT 960
Db 2311 CTCAAGCTGAGCTTTTGTGGCCATCCACCACTGCTGTTTGGCCACCGTCAAGACCCGCT 2370
Qy 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCCACTGTCTCTGATCCAGCGAGGT 1020
Db 2371 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCCACTGTCTCTGATCCAGCGAGGT 2430
Qy 1021 ACCAATGGCACTCCCGATCAGGCTAAAGCTTTGCCAATGTTTCTGATGGCTTAAGTGCC 1080
Db 2431 ACCAATGGCACTCCCGATCAGGCTAAAGCTTTGCCAATGTTTCTGATGGCTTAAGTGCC 2490
Qy 1081 TGGGTTTGTCTTAATAGAACTGAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db 2491 TGGGTTTGTCTTAATAGAACTGAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2550
Qy 1141 CCAGGCTTCTTAATAGAACTGAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Db 2551 CCAGGCTTCTTAATAGAACTGAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2609
Qy 1201 TCTGTGAGGCGCAAGAACCCAGGCTCAGAGAAAGTGGAGCTTGGCCACCAATTTGGGAAG 1257
Db 2610 TCCATAAGGCGCAAGAACCCAGGCTCAGAGAAAGTGGAGCTTGGCCACCAATTTGGGAAG 2666

```

Job time : 5585.86 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2005, 11:11:14 ; Search time 609.913 Seconds
(without alignments)
11438.492 Million cell updates/sec

Title: US-09-319-156B-12
Perfect score: 1329
Sequence: 1 tcaaatcgaagagcttag.....gcaagatccccagtaaca 1329

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_23Sep04:*

1:	Geneseqn1980s:*
2:	Geneseqn1930s:*
3:	Geneseqn2000s:*
4:	Geneseqn2001as:*
5:	Geneseqn2001bs:*
6:	Geneseqn2002as:*
7:	Geneseqn2002bs:*
8:	Geneseqn2003as:*
9:	Geneseqn2003bs:*
10:	Geneseqn2003cs:*
11:	Geneseqn2003ds:*
12:	Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1326	99.8	1329	2	AAV43219 Multiple
2	1326	99.8	1329	2	Aax25704 Clone SM6
3	1326	99.8	1329	12	Adg14849 MSRV asso
4	1299	97.7	1329	9	Adb84403 MSRV-1 as
5	1035.2	77.9	56093	6	Ab161744 Colon ade
6	1029.4	77.5	9502	10	Adf59718 Human con
7	1023.2	77.0	10499	3	Abn97929 Human ret
8	801.6	60.3	7466	5	Aas68626 DNA encod
9	778.8	58.6	1393	4	Aas31002 Human dia
10	775.4	58.3	2030	3	AAAG63826
11	766.8	57.7	2074	6	AAD41225 Human EMB
12	747.6	56.3	2046	8	ACC46747 Human dit
13	731.8	55.1	2946	2	Aax77526 Human sec
14	731.8	55.1	2946	3	Aaz59468 Human sec
15	731.8	55.1	2946	10	ADC38776 Human cDN
16	730.8	55.0	2930	6	AAD24195 Human syn
17	730.2	54.9	2781	5	AAF55630 Nucleotid
18	728.6	54.8	6394	5	AAS84210 DNA encod
19	722.6	54.4	7582	2	AAX25665 Comple
20	722.6	54.4	7582	3	AAAS5215 Human end
21	721.2	54.3	1136	2	AAX25660 Human end

22	721.2	54.3	1136	3	AAAS5210
23	712.6	53.6	2782	2	AAX25661
24	712.6	53.6	2782	3	AAAS5211
25	712.6	53.6	2782	5	AAH20069
26	704.8	53.0	1894	4	AAI14608
27	704.8	53.0	1894	4	ABA56337
28	704.8	53.0	1894	4	AAI35980
29	704.8	53.0	1894	4	ABA45822
30	704.8	53.0	1894	4	ABA25978
31	704.8	53.0	1894	4	AAK30018
32	704.8	53.0	1894	4	AAK04516
33	704.8	53.0	1894	4	ABS29670
34	704.8	53.0	1894	5	AAI04422
35	704.8	53.0	1894	6	ABS04589
36	680.6	51.2	2942	5	AAAS77313
37	651	49.0	893	5	AAS65964
38	643.4	48.4	46340	3	ABN97978
39	639.8	48.1	783	2	AAX25669
40	610.4	45.9	3464	10	ADR09587
41	607	45.7	2784	3	ABN97930
C 42	601.4	45.3	157090	12	ADO47194
43	596.8	44.9	711	3	ABN97947
C 44	593.6	44.7	2316	9	ACC84313
45	580.6	43.7	711	3	ABN97946

ALIGNMENTS

RESULT 1
AAV43219
ID AAV43219 standard; cDNA; 1329 BP.
XX
AC AAV43219;
XX
DT 29-DEC-1998 (first entry)
XX
DE Multiple sclerosis associated retrovirus fragment 6.
DE Multiple sclerosis associated retrovirus; MSRV; MS; pol gene; gag gene;
KW env gene; rheumatoid arthritis-associated virus; ss.
XX
OS Multiple sclerosis associated retrovirus.
XX
FH Location/Qualifiers
FT CDS
FT 2..490
FT /tag= a
FT /product= "Encodes protein AAV71069"
FT /transl_except= (pos:77-79, appears to code for a stop
FT codon)
FT /transl_except= (pos:125-127, appears to code for a stop
FT codon)
FT /transl_except= (pos:137-139, appears to code for a stop
FT codon)
XX
WO9823755-A1.
PD 04-JUN-1998.
XX
PF 26-NOV-1997; 97WO-IB001482.
XX
PR 26-NOV-1996; 96US-00756429.
XX
PA (INMR) BIO MERIEUX.
XX
PI Perron H, Beseme F, Bedin F, Paranhos-Baccala G;
PI Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B;
DR WPI; 1998-322732/28.
DR P-PSDB; AAV71069.
XX
PT New nucleic acid from retroviruses - useful for diagnosis, prevention and
PT treatment of, e.g. multiple sclerosis.

XX PS Disclosure; Page 187-188; 286pp; English.

XX CC The present sequence represents a multiple sclerosis (MS) associated

CC retrovirus (MSRV) genomic fragment used in the method of the invention.

CC CC The invention provides complete or partial genomic sequences of the MSRV-

CC 1 pol gene, gag gene and env gene, and polypeptides encoded by these

CC genes. The invention also provides antibodies raised against the

CC polypeptides. The genomic sequences, polypeptides and antibodies are also

CC claimed useful for diagnosing infection by MS and rheumatoid arthritis-

CC associated viruses, and also for prevention and treatment of infection

XX with these viruses

SQ Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 0 U; 3 Other;

Query Match 99.8%; Score 1326; DB 2; Length 1329;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCAAAAGAGGGGAACTGTTTATTTT 60

DB 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCAAAAGAGGGGAACTGTTTATTTT 60

QY 61 AGGGAAGAATGCTGTAGTATGTTAATCAATCTGGAATCATTTACTGAGAAAGTTAAGA 120

DB 61 AGGGAAGAATGCTGTAGTATGTTAATCAATCTGGAATCATTTACTGAGAAAGTTAAGA 120

QY 121 AATTGTAGATCGAATAAATGATGAGCAGAGGACCTTCAAAAACACTGACCCCTGGGCGCT 180

DB 121 AATTGTAGATCGAATAAATGATGAGCAGAGGACCTTCAAAAACACTGACCCCTGGGCGCT 180

QY 181 CCTCAGCAATGGATGCCCTGGACCTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATTT 240

DB 181 CCTCAGCAATGGATGCCCTGGACCTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATTT 240

QY 241 TTTACTCTCTTTGGACCCCTGATCTTCAACTCTCTTGTAAAGTTTGTCTCTTCAGAAAT 300

DB 241 TTTACTCTCTTTGGACCCCTGATCTTCAACTCTCTTGTAAAGTTTGTCTCTTCAGAAAT 300

QY 301 TGAAGCTGTAAAGCTACAAATAGTTTCTTCAAAATGGAACCCAGATGCAATGCAATGCTAA 360

DB 301 TGAAGCTGTAAAGCTACAAATAGTTTCTTCAAAATGGAACCCAGATGCAATGCAATGCTAA 360

QY 361 AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGTTAATGACATTTGA 420

DB 361 AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGTTAATGACATTTGA 420

QY 421 AGTCACCCCTCCGAGGAAATCTCAACTGCAACACCCCTACTACCTCCTCAATTCAGTAGG 480

DB 421 AGTCACCCCTCCGAGGAAATCTCAACTGCAACACCCCTACTACCTCCTCAATTCAGTAGG 480

QY 481 AAGCAGTTAGCAGAGTTGTGAGCAACCTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 540

DB 481 AAGCAGTTAGCAGAGTTGTGAGCAACCTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 540

QY 541 GGGTGGACTGAGACAGGACTAGCTGGAATTTCTAGGCTGACTAAGAATCCCNAAAGCCT 600

DB 541 GGGTGGACTGAGACAGGACTAGCTGGAATTTCTAGGCTGACTAAGAATCCCNAAAGCCT 600

QY 601 ANCTGGGAAGGTGACCGCATCTCTTTAAACATGGGGCTTGGCACTTAGCTCACACCCG 660

DB 601 ANCTGGGAAGGTGACCGCATCTCTTTAAACATGGGGCTTGGCACTTAGCTCACACCCG 660

QY 661 ACCAATCAGAGCTCACTAAATGCTTAATCAGCAAAAAACAGAGGTAAGCAATAGCC 720

DB 661 ACCAATCAGAGCTCACTAAATGCTTAATCAGCAAAAAACAGAGGTAAGCAATAGCC 720

QY 721 AATCATCTATTGCTGAGACAGCAGCGGGAAGGACAGGATTGGGATATAAATCTCAGGCA 780

DB 721 AATCATCTATTGCTGAGACAGCAGCGGGAAGGACAGGATTGGGATATAAATCTCAGGCA 780

QY 781 TTCAGCAGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGATGGGAGCTCTGTTTT 840

DB 781 TTCAGCAGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGATGGGAGCTCTGTTTT 840

QY 841 CACTCTATTCTACTCTATTAAATCATGCAATGCACTCTTCTGGTCCGTTTATGG 900

DB 841 CACTCTATTCTACTCTATTAAATCATGCAATGCACTCTTCTGGTCCGTTTATGG 900

QY 901 CTCAAGCTGAGCTTTTGTTCGGCATCCACACTGCTGTTTGGCACCGTACAGACCGCT 960

DB 901 CTCAAGCTGAGCTTTTGTTCGGCATCCACACTGCTGTTTGGCACCGTACAGACCGCT 960

QY 961 GCTGACTTCCATCCCTTTTGGATCCAGCAGAGTGTCCACTGTCTCTGATCCAGCGAGT 1020

DB 961 GCTGACTTCCATCCCTTTTGGATCCAGCAGAGTGTCCACTGTCTCTGATCCAGCGAGT 1020

QY 1021 ACCCATTTGCCACTCTCCGATCAGGCTAAAGGCTTGCCATTTCTCTGCAATGGCTAAGTGCC 1080

DB 1021 ACCCATTTGCCACTCTCCGATCAGGCTAAAGGCTTGCCATTTCTCTGCAATGGCTAAGTGCC 1080

QY 1081 TGGGTTTGTCTTAATAGAACTGAACTGCTGGTCACTGGGTTCCATGTTCTCTTCCATGAC 1140

DB 1081 TGGGTTTGTCTTAATAGAACTGAACTGCTGGTCACTGGGTTCCATGTTCTCTTCCATGAC 1140

QY 1141 CCACGGCTTCTAATAGAGCTATAAACACTCACCGCATGGCCCAAGATTCCATTCCTTTGGTA 1200

DB 1141 CCACGGCTTCTAATAGAGCTATAAACACTCACCGCATGGCCCAAGATTCCATTCCTTTGGTA 1200

QY 1201 TCTGTGAGGCCAAGAACCCCGAGTCAAGAAANGTAGGGCTTGCCACATTTGGGAAAGTGG 1260

DB 1201 TCTGTGAGGCCAAGAACCCCGAGTCAAGAAANGTAGGGCTTGCCACATTTGGGAAAGTGG 1260

QY 1261 CCCACTGGCCATTTTGTAGGGCCCAACACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320

DB 1261 CCCACTGGCCATTTTGTAGGGCCCAACACCATCTTGGGAGCTGTGGGAGCAAGGATCCC 1320

QY 1321 CCAGTAACA 1329

DB 1321 CCAGTAACA 1329

RESULT 2

AAAX29704

ID AAAX29704 standard; DNA; 1329 BP.

XX AC AAAX29704;

XX AC

DT 17-OCT-2003 (revised)

DT 27-AUG-2003 (revised)

DT 08-JUN-1999 (first entry)

XX DE

DE Clone SM6 from MSRV-1.

XX Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;

KW rheumatoid polyarthritis; ss.

XX Viruses.

XX FR2765588-A1.

PN

PD 08-JAN-1999.

XX

PF 07-JUL-1997; 97FR-00008816.

XX

PR 07-JUL-1997; 97FR-00008816.

XX

PA (INMR) BIO MERIEUX.

XX

DR WPI; 1999-098275/09.

XX

DR P-PSDB; AAW99554.

XX

PT Nucleic acid sequences of retrovirus called MSRV-1 - associated with

PT multiple sclerosis or rheumatoid polyarthritis.

XX

PS Claim 1; Page 39-40; 83pp; French.

XX This sequence represents clone 5M6 from a novel multiple sclerosis
CC related virus type 1 (MSRV). The sequence can be used in diagnostic,
CC prophylactic or therapeutic compositions to inhibit expression of a
CC multiple sclerosis related virus and/or virus associated with rheumatoid
CC polyarthritis. (Updated on 27-AUG-2003 to correct OS field.) (Updated on
CC 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 0 U; 3 Other;

Query Match 99.8%; Score 1326; DB 2; Length 1329;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAAATCGAAGAGCTTTAGACTTGTCTTAACCGGCAAAAGAGGGGGAACCTGTTTATTTT 60
DB 1 TCAAAATCGAAGAGCTTTAGACTTGTCTTAACCGGCAAAAGAGGGGGAACCTGTTTATTTT 60

QY 61 AGGGGAAGATGCTGTAGTATGTTAATCAATCTGGAATCAATCTGAGAAAGTTAAAGA 120
DB 61 AGGGGAAGATGCTGTAGTATGTTAATCAATCTGGAATCAATCTGAGAAAGTTAAAGA 120

QY 121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCAACCTGGGGCCT 180
DB 121 AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCAACCTGGGGCCT 180

QY 181 CCTCAGCCAAATGATGCCCTGGACTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATAT 240
DB 181 CCTCAGCCAAATGATGCCCTGGACTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATAT 240

QY 241 TTTACTCCTCTTTGGACCTGTATCTTCAACTCTCTTGTAGTTTGTCTCTCCAGAT 300
DB 241 TTTACTCCTCTTTGGACCTGTATCTTCAACTCTCTTGTAGTTTGTCTCTCCAGAT 300

QY 301 TGAAGCTGTAAAGCTCAAAATAGTTCTTCAATGGAACCCAGATGCAGTCCATGACTAA 360
DB 301 TGAAGCTGTAAAGCTCAAAATAGTTCTTCAATGGAACCCAGATGCAGTCCATGACTAA 360

QY 361 AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGTTAAATGACATTGA 420
DB 361 AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGTTAAATGACATTGA 420

QY 421 AGTCACCCCTCCGAGGAATCTCAATCTGACAGAACCCCTACTACACTCCAATTCAGTAG 480
DB 421 AGTCACCCCTCCGAGGAATCTCAATCTGACAGAACCCCTACTACACTCCAATTCAGTAG 480

QY 481 AAGCAGTTAGAGCAGTTGTGAGCAACCTCCCAACAGACTTTGGGTTTCTCTGTTGAGA 540
DB 481 AAGCAGTTAGAGCAGTTGTGAGCAACCTCCCAACAGACTTTGGGTTTCTCTGTTGAGA 540

QY 541 GGGTGGACTGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAGCCT 600
DB 541 GGGTGGACTGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAGCCT 600

QY 601 ANCTGGGAAGGTGACCGCATTCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
DB 601 ANCTGGGAAGGTGACCGCATTCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660

QY 661 ACCAATCAGAGAGCTCAGTAAATGCTAATCAGCAAAAGAGGTAAGCAATAGCC 720
DB 661 ACCAATCAGAGAGCTCAGTAAATGCTAATCAGCAAAAGAGGTAAGCAATAGCC 720

QY 721 AATCATCTATTGCTGAGACAGCGGGAAGGACAGGATTGGGATATAAACTCAGGCA 780
DB 721 AATCATCTATTGCTGAGACAGCGGGAAGGACAGGATTGGGATATAAACTCAGGCA 780

QY 781 TTCAAGCAGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGTATGGAGCTCTGTTT 840
DB 781 TTCAAGCAGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGTATGGAGCTCTGTTT 840

QY 841 CACTCTATTCTACTTATTAAATCATGCACTCTCTGTCCTGTCGTTTATATGG 900
DB 841 CACTCTATTCTACTTATTAAATCATGCACTCTCTGTCCTGTCGTTTATATGG 900

QY 901 CTCAGCTGAGCTTTTGTTCGCATCCACCACTGCTGTTTGCACCGTCCACAGACCGCT 960
DB 901 CTCAGCTGAGCTTTTGTTCGCATCCACCACTGCTGTTTGCACCGTCCACAGACCGCT 960

QY 961 GCTGACTTCCATCCCTTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020
DB 961 GCTGACTTCCATCCCTTTTGGATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCGAGGT 1020

QY 1021 ACCCATTCGCACTCCCGATCAGGCTAAAGCTTGGCAATTTGCTCTCATGGCTTAAGTCC 1080
DB 1021 ACCCATTCGCACTCCCGATCAGGCTAAAGCTTGGCAATTTGCTCTCATGGCTTAAGTCC 1080

QY 1081 TGGGTTTGTCTTAATAGAACTGAACCTGCTCACTGGGTTCATGTTCTCTTCCATGAC 1140
DB 1081 TGGGTTTGTCTTAATAGAACTGAACCTGCTCACTGGGTTCATGTTCTCTTCCATGAC 1140

QY 1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCTCTTGGTA 1200
DB 1141 CCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCCATTCTCTTGGTA 1200

QY 1201 TCTGTAGGCCCAAGAACCCAGGTCCAGAAAGTGGCTTGCACCATTTGGGAAGTGG 1260
DB 1201 TCTGTAGGCCCAAGAACCCAGGTCCAGAAAGTGGCTTGCACCATTTGGGAAGTGG 1260

QY 1261 CCCACTGCAATTTTGTAGCGGCCCAACCACTCTTGGAGCTGTGGAGCAAGGATCCC 1320
DB 1261 CCCACTGCAATTTTGTAGCGGCCCAACCACTCTTGGAGCTGTGGAGCAAGGATCCC 1320

QY 1321 CCAGTAACA 1329
DB 1321 CCAGTAACA 1329

RESULT 3
ADG14849
ID ADG14849 standard; cdna; 1329 BP.
XX
AC ADG14849;
XX
DT 26-FEB-2004 (first entry)
XX
DE MSRV associated cDNA #11.
XX
KW ss; pol gene; retrovirus; multiple sclerosis; rheumatoid arthritis.
XX OS Multiple sclerosis associated retrovirus.
XX PN US2003198647-A1.
XX
PD 23-OCT-2003.
XX
PF 03-APR-2002; 2002US-00114104.
XX
PR 26-NOV-1996; 96US-00756429.
XX 26-NOV-1997; 97US-00979847.
XX
PA (INMR) BIO MERIEUX.
XX
PI Perron H, Beseme F, Bedin F, Paranhos-Baccala G;
PI Komurian-Pradel F, Jollivet-Reynaud C, Mandrand B, Garson JA, Tuke PW;
XX WPI; 2004-032461/03.
XX
PT New isolated nucleic acid and their fragments having the pol gene of a
PT retrovirus, useful for diagnosing, preventing and/or treating multiple
PT sclerosis and/or rheumatoid arthritis.
XX
PS Disclosure; SEQ ID NO 108; 193pp; English.
XX
CC The invention relates to an isolated nucleic acid which comprises the pol
CC gene of a retrovirus associated with multiple sclerosis or rheumatoid
CC arthritis. The methods and compositions of the present invention are

CC useful for diagnosing, preventing and/or treating multiple sclerosis
CC and/or rheumatoid arthritis. The present sequence is used in the
XX exemplification of the invention.

SQ Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 0 U; 3 Other;

Query Match 99.8%; Score 1326; DB 12; Length 1329;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TCAAAATCGAAGAGCTTACGATTCCTAACCGCAAGAGGGGAGACCTGTATATTTT 60
Dd 1 TCAAAATCGAAGAGCTTACGATTCCTAACCGCAAGAGGGGAGACCTGTATATTTT 60
QY 61 AGGGGAAGAAATGCTGTAGTATGTTAATCAATCTGGAATCATTTACTGAGAAAGTTAAGA 120
Dd 61 AGGGGAAGAAATGCTGTAGTATGTTAATCAATCTGGAATCATTTACTGAGAAAGTTAAGA 120
QY 121 AATTTGAGATCGAATATAATGTAGAGCAGAGACCTTCAAAAACATGACCCCTGGGCGCT 180
Dd 121 AATTTGAGATCGAATATAATGTAGAGCAGAGACCTTCAAAAACATGACCCCTGGGCGCT 180
QY 181 CCTCAGCAATGATGATGCTGGAATCTCCCTTCTTAGGACCTCTAGCAGCTATAATTT 240
Dd 181 CCTCAGCAATGATGATGCTGGAATCTCCCTTCTTAGGACCTCTAGCAGCTATAATTT 240
QY 241 TTTTACTCCTCTTTGGACCCCTGTATCTTCAACTTCTTGTAAAGTTTGTCTCTTCCAGAA 300
Dd 241 TTTTACTCCTCTTTGGACCCCTGTATCTTCAACTTCTTGTAAAGTTTGTCTCTTCCAGAA 300
QY 301 TGAAGCTGTAAGAGTCAAAATAGTTCTTCAATTTGGAACCCAGATGAGTCCATGACTAA 360
Dd 301 TGAAGCTGTAAGAGTCAAAATAGTTCTTCAATTTGGAACCCAGATGAGTCCATGACTAA 360
QY 361 AATCTACCGTGGACCCCTGACCGGCTGCTAGACTATGCTCTGATGTTAATGACATTTGA 420
Dd 361 AATCTACCGTGGACCCCTGACCGGCTGCTAGACTATGCTCTGATGTTAATGACATTTGA 420
QY 421 AGTCACCCCTCCCGAGGAAATCTCACTGCAACACCCCTACTACACTCCAAATTCAGTAGG 480
Dd 421 AGTCACCCCTCCCGAGGAAATCTCACTGCAACACCCCTACTACACTCCAAATTCAGTAGG 480
QY 481 AAGCAGTTAGAGAGTTGTAGCCAACTCCCAACAGTACTTGGGTTTCTGTGTTGAGA 540
Dd 481 AAGCAGTTAGAGAGTTGTAGCCAACTCCCAACAGTACTTGGGTTTCTGTGTTGAGA 540
QY 541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCCAAGCCT 600
Dd 541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCCAAGCCT 600
QY 601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
Dd 601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
QY 661 ACCAATCAGAGAGTCACTAAATGCTAATCAGGCAAAAACAGGAGTTAAAGCAATAGCC 720
Dd 661 ACCAATCAGAGAGTCACTAAATGCTAATCAGGCAAAAACAGGAGTTAAAGCAATAGCC 720
QY 721 AATCATCTATTGCTGAGAGCAGCGGGAAGCAGCAAGGATTGGGATATAAATCTCAGGCA 780
Dd 721 AATCATCTATTGCTGAGAGCAGCGGGAAGCAGCAAGGATTGGGATATAAATCTCAGGCA 780
QY 781 TTCAAGCCAGCAACAGCAACCCCTTTGGTCCCTCCCTCCCATTTGTATGGGAGCTCTGTTT 840
Dd 781 TTCAAGCCAGCAACAGCAACCCCTTTGGTCCCTCCCTCCCATTTGTATGGGAGCTCTGTTT 840
QY 841 CACTCTATTCTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTTTATGAG 900
Dd 841 CACTCTATTCTCTATTAAATCATGCAACTGCACTCTTCTGGTCCGTTTATGAG 900
QY 901 CTCAGCTAGCTTTGTTGCGCATCCACCACTGCTTTTGGCCATCCAGACCCGCT 960
Dd 901 CTCAGCTAGCTTTGTTGCGCATCCACCACTGCTTTTGGCCATCCAGACCCGCT 960
```

RESULT 4

ADB84403

ID ADB84403 standard; DNA; 1329 BP.

XX AC ADB84403;

XX DT 04-DEC-2003 (first entry)

XX MSRV-1 associated DNA sequence #21.

XX ds; multiple sclerosis; rheumatoid arthritis; gag; pol;

XX reverse transcriptase; ribonuclease H.

XX Unidentified;.

XX US2003039664-A1.

XX 27-FEB-2003.

XX 26-NOV-1997; 97US-00979847.

XX 26-NOV-1996; 96US-00756429.

XX (PERK/) PERRON H.

XX (BESE/) BESEME F.

XX (BEDI/) BEDIN F.

XX (PARA/) PARANHOS-BACCALA G.

XX (KOMU/) KOMURIAN-PRADEL F.

XX (JOLI/) JOLIVET-REYNAUD C.

XX (MAND/) MANDRAND B.

XX (GARS/) GARSON J A.

XX (TUKU/) TUKU P W.

XX Perron H, Beseme F, Bedin F, Paranhos-Baccala G;

XX Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B, Garson JA, Tuke PW;

XX WPI; 2003-512253/48.

XX New isolated or purified nucleic acid associated with multiple sclerosis

XX and/or rheumatoid arthritis, useful for detecting a virus associated with

XX multiple sclerosis or rheumatoid arthritis in a biological sample.

XX Claim 31; Page 80; 193pp; English.

XX The invention relates to an isolated or purified nucleic acid from a
CC virus associated with multiple sclerosis and/or rheumatoid arthritis,
CC multiple sclerosis-associated virus (MSRV)-1. The nucleic acids comprise
CC pol, gag or reverse transcriptase genes (or their fragments) encoding the
CC proteins or defined peptides (including immunodominant peptides,
CC antigenic peptides or conserved motifs). Also included are a process for
CC detecting a virus associated with multiple sclerosis or rheumatoid
CC arthritis in a biological sample, a nucleic acid probe for the detection
CC of a virus associated with multiple sclerosis or rheumatoid arthritis, a
CC primer for the amplification by polymerisation of a nucleic acid of a
CC viral material associated with multiple sclerosis or rheumatoid
CC arthritis, a polypeptide exhibiting an inhibitory activity on the
CC proteolytic, reverse transcriptase or ribonuclease H activity from MSRV,
CC and an antibody directed against the MSRV-1 virus obtained by
CC immunogenically reacting a human or animal body or cells with an
CC immunologic agent consisting of the antigenic polypeptide defined above.
CC The nucleic acids are useful for detecting a biological sample a virus
CC associated with multiple sclerosis or rheumatoid arthritis, or for
CC detecting in a biological sample, the presence of or exposure to a virus
CC associated with multiple sclerosis or rheumatoid arthritis. The present
CC sequence is a claimed MSRV-associated sequence whose identity cannot be
CC accurately determined. Note: The SEQ ID numbers for the sequences as
CC displayed in the main body of the patent do not match the SEQ ID numbers
CC in the sequence listing. Consequently those sequences mentioned in the
CC claims may not be the sequences the authors intended to claim.
XX
SQ Sequence 1329 BP; 339 A; 346 C; 282 G; 332 T; 0 U; 30 Other;

Query Match 97.78; Score 1299; DB 9; Length 1329;
Best Local Similarity 98.04; Pred. No. 0;
Matches 1302; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY	1	TCAAAATCGAAGAGCTTTAGACTTGTCAACGCGCAAGAGGGGAACTGTTTATTTT	60
DB	1	TCAAAATCGAAGAGCTTTAGACTTGTCAACGCGCAAGAGGGGAACTGTTTATTTT	60
QY	61	AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA	120
DB	61	AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGN	120
QY	121	AATTTGAGATCGAATAATATGTAGAGCAGAGGACCTTCAAAACACTGACACCTGGGGCCT	180
DB	121	AATTTGAGATCGAATAATATGTAGAGCAGAGGACCTTCAAAACACTGACACCTGGGGCCN	180
QY	181	CCTCAGCAATGATGATGCTGCTCCCTCTCTTAGGACCTCTAGCAGCTATAATTT	240
DB	181	CCTCAGCAATGATGATGCTGCTCCCTCTCTTAGGACCTCTAGCAGCTATAATATN	240
QY	241	TTTACTCTCTTTGGACCCCTGTATCTTCAACTTCCTTGTAAAGTTTGTCTCTTCCAGAA	300
DB	241	TTTACTCTCTTTGGACCCCTGTATCTTCAACTTCCTTGTAAAGTTTGTCTCTTCCAGAA	300
QY	301	TGAAGCTGTAAGACTAATAATAGTTCTTCAAAATGGAAACCCAGATGAGTCCATGACTAA	360
DB	301	TGAAGCTGTAAGACTAATAATAGTTCTTCAAAATGGAAACCCAGATGAGTCCATGACTAN	360
QY	361	AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTGTAGTGTAAATGACATTGA	420
DB	361	AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTGTAGTGTAAATGACATTGN	420
QY	421	AGTCACCCCTCCGAGGAATCTCAACTGCAACACCCCTACTACACTCCCAATTCAGTAGG	480
DB	421	AGTCACCCCTCCGAGGAATCTCAACTGCAACACCCCTACTACACTCCCAATTCAGTAGN	480
QY	481	AAGCAGTTAGACGAGTTGTGAGCCAACTCCCAACAGTACTTGGGTTTCTGTTGAGA	540
DB	481	AAGCAGTTAGACGAGTTGTGAGCCAACTCCCAACAGTACTTGGGTTTCTGTTGAGN	540
QY	541	GGGTGGACTGAGACGAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAGCCT	600
DB	541	GGGTGGACTGAGACGAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCNAGCCN	600

601 ANCTGGGAGGTGACCGCATCCATCTTTAAACATGGGCTTGCAACTTAGCTCACACCG 660
601 ANCTGGGAGGTGACCGCATCCATCTTTAAACATGGGCTTGCAACTTAGCTCACACCN 660
661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAACAGAGGTAAGCAATAGCC 720
661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAACAGAGGTAAGCAATAGCN 720
721 AATCATCTATTGCTGAGACAGCGGGAAGCAAGGATGGGATATAAACTCAGGCA 780
721 AATCATCTATTGCTGAGACAGCGGGAAGCAAGGATGGGATATAAACTCAGGCA 780
781 TTCAAGCCGACACAGCAACCCCTTTGGTCCCTCCCATCTCTCTGGTCCGTTGTTT 840
781 TTCAAGCCGACACAGCAACCCCTTTGGTCCCTCCCATCTCTCTGGTCCGTTGTTT 840
841 CACTCTATTCTACTCTATTAAATCATGCAATGCACTCTCTCTGGTCCGTTGTTT 900
841 CACTCTATTCTACTCTATTAAATCATGCAATGCACTCTCTCTGGTCCGTTGTTT 900
901 CTCAGCTGAGCTTTTGTGCGCATCCACACTGCTGTTTGGCCCGTCAAGACCCGCT 960
901 CTCAGCTGAGCTTTTGTGCGCATCCACACTGCTGTTTGGCCCGTCAAGACCCGCT 960
961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTCTCTGATCCAGCGAGT 1020
961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTCTCTGATCCAGCGAGN 1020
1021 ACCATTGGCCTCCGATCAGCTAAAGCTTTGCCATTTGCTCTGATGGCTTAAGTGCC 1080
1021 ACCATTGGCCTCCGATCAGCTAAAGCTTTGCCATTTGCTCTGATGGCTTAAGTGNN 1080
1081 TGGTTTGTCTTAATAGAACTGAACACTGTGTCACTGGGTTTCCATGGTTCTCTTCCATGAC 1140
1081 TGGTTTGTCTTAATAGAACTGAACACTGTGTCACTGGGTTTCCATGGTTCTCTTCCATGNN 1140
1141 CACGGCTTCTTAATAGAGCTATAAACACTCAACCCATGGCCCAAGATTCATCTCTTGGTA 1200
1141 CACGGCTTCTTAATAGAGCTATAAACACTCAACCCATGGCCCAAGATTCATCTCTTGGNN 1200
1201 TCTGTGAGCCCAAGAACCCAGGTGAGAACTGAGGCTTGCCACCATTTGGGAGTGG 1260
1201 TCTGTGAGCCCAAGAACCCAGGTGAGAACTGAGGCTTGCCACCATTTGGGAGATNN 1260
1261 CCACCTGCCATTTTGTAGCGGCCACCACTCTTGGAGCTTGTGGAGCAAGGATCCC 1320
1261 CCACCTGCCATTTTGTAGCGGCCACCACTCTTGGAGCTTGTGGAGCAAGGATCNN 1320
1321 CCAGTAACA 1329
1321 CCAGTAACA 1329

RESULT 5
ABL61744
ID ABL61744 standard; DNA; 56093 BP.
XX
AC ABL61744;
XX
XX 15-MAY-2002 (first entry)
XX
XX Colon adenocarcinoma related gene sequence SEQ ID NO:81.
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
XX cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
XX gene; ds.
XX
XX Homo sapiens.
XX
XX WO200194629-A2.
XX
XX 13-DEC-2001.

XX 30-MAY-2001; 2001WO-US010838.
XX
PR 05-JUN-2000; 2000US-0209473P.
PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0233133P.
PR 18-SEP-2000; 2000US-0233133P.
PR 20-SEP-2000; 2000US-0233617P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235280P.
PR 25-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237295P.
PR 02-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX
XX Claim 1; SEQ ID NO 81; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 847 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell

Qy	961	GCTGACTTCCATCCCTTTGGATCCACGACGAGTGTCCACTGTGCTCCTGATCCAGCGAGT	1020
Db	37966	GCTGACTTCCCATCCCTCTGGATCATGACAGGGTGTCCGTGTGCTCTGATCCAGCGAGGC	38025
Qy	1021	ACCATTGGCACATCCCGATCAGGCTAAAGGCTTGCATGTTCCCTGCGATGGCTAAGTGCC	1080
Db	38026	ACCATTGGCGCTCCCAATCGGGCTTAAGGCTTGCATGTTCCCTGCGATGGCTAAGTGCC	38085
Qy	1081	TGGGTTGTCTCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCCTCTTCCATGAC	1140
Db	38086	TGGGTTCACTCTAATTTGAGCTGAACTAGTACTCTGGGTTCCATGGTTCCTCTTCTGTGAC	38145
Qy	1141	CCACGGCTTCTAATAGAGCTATAAACACTCACCGCATGGCCCCAAGATTCCATTCCTTTGGTA	1200
Db	38146	CCACAGCTTCTAATAGAGCTATAAACACTCACCGCATGGCCCCAAGGTTCCATTCCTTT-GAA	38204
Qy	1201	TCTGTGAGGCCAGAACCCCGCTCAGAGAAAGTGAAGCTTCCGACCATTTGGGGAAG	1257
Db	38205	TCCATAAGGCCAAGAACCCCGCTCAGAGAAACGAGGCTTGCCACCATCTTGGGAG	38261
RESULT 6			
ADF59718			
ID	ADF59718 standard; cDNA; 9502 BP.		
XX	AC	AC	XX
XX	ADF59718;		
DT	12-FEB-2004 (first entry)		
XX	Human contig polynucleotide sequence SEQ ID NO:2085.		
DE	biological activity; genetic engineering; hybridisation probe; oligomer;		
KW	primer; chromosome mapping; gene mapping; recombinant protein production;		
KW	human; gene; ss.		
XX	Homo sapiens.		
XX	WO2003080795-A2.		
XX	02-OCT-2003.		
XX	09-AUG-2002; 2002WO-US025485.		
PF	09-AUG-2001; 2001US-0311261P.		
PR	(HYSE-) HYSEQ INC.		
PA	Tang YT, Yang Y, Wang Z, Weng G, Ma Y;		
XX	WPI; 2003-876918/81.		
DR	P-PSDB; ADF60170.		
XX	New polynucleotides, useful as hybridization probes, oligomers or		
PT	primers, for chromosome or gene mapping, for the recombinant production		
PT	of proteins, and for generating antisense DNA or RNA.		
XX	Example 2; SEQ ID NO 2085; 571pp; English.		
XX	The present invention describes isolated polynucleotide sequences (I),		
CC	which encode polypeptides (II) with biological activity. Also described:		
CC	(1) a vector comprising (1); (2) an expression vector comprising (1); (3)		
CC	a host cell genetically engineered to comprise (I) which is operatively		
CC	associated with a regulatory sequence that modulates expression of (I) in		
CC	the host cell; (4) a polypeptide (II) encoded by (1); (5) a composition		
CC	comprising the polypeptide of (4) and a carrier; (6) an antibody directed		
CC	against the polypeptide of (4); (7) detecting (I) or the polypeptide of		
CC	(4) in a sample; (8) identifying a compound that binds to the polypeptide		
CC	of (4); (9) producing the polypeptide of (4); and (10) a collection of		
CC	polynucleotides comprising at least one of the polynucleotide sequences		
CC	(I). The polynucleotides (I) can be used as hybridisation probes,		
CC	oligomers or primers, for chromosome or gene mapping, for the recombinant		
CC	production of proteins, and for generating antisense DNA or RNA. The		

Db 4330 TAAACAGAGCTATAACACTACCGCATGGCCCAAGGTTCCGTTCTTGTATTCGTGAGGC 4389
QY 1211 CAAGAACCCAGGTCAGAGANGTGGAGCTTGGCCACCATTTGGGAAGTGGCCCACTGCCA 1270
Db 4390 CAAGAACCCAGGTCAGAGAACAAAGGCTTGGACCATTTGGAGGCGCTGTACCG 4449
QY 1271 TTTTGGTAGCGGCCCAACCATCTTTGGGAGCTGTGGGACCAAGGATCCCCAGTAGTAA 1329
Db 4450 TCTTGGAGTGGTTTCAACCATCTTTGGGAGCTCTGTGACCAAGGACCCCGGTAA 4508

RESULT 9

AA31002
ID AAS31002 standard; cDNA; 1393 BP.

AC AAS31002;

XX 04-DEC-2001 (first entry)

XX Human diagnostic and therapeutic polynucleotide (DITHP) #17.

XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
KW respiratory disorder; ss.

OS Homo sapiens.

XX WO200162927-A2.

PN 30-AUG-2001.

XX 21-FEB-2001; 2001WO-US006059.

XX 24-FEB-2000; 2000US-0184693P.

PR 24-FEB-2000; 2000US-0184693P.

PR 24-FEB-2000; 2000US-0184698P.

PR 24-FEB-2000; 2000US-0184768P.

PR 24-FEB-2000; 2000US-0184769P.

PR 24-FEB-2000; 2000US-0184770P.

PR 24-FEB-2000; 2000US-0184771P.

PR 24-FEB-2000; 2000US-0184772P.

PR 24-FEB-2000; 2000US-0184773P.

PR 24-FEB-2000; 2000US-0184774P.

PR 24-FEB-2000; 2000US-0184776P.

PR 24-FEB-2000; 2000US-0184777P.

PR 24-FEB-2000; 2000US-0184797P.

PR 24-FEB-2000; 2000US-0184813P.

PR 24-FEB-2000; 2000US-0184837P.

PR 24-FEB-2000; 2000US-0184841P.

PR 24-FEB-2000; 2000US-0185213P.

PR 24-FEB-2000; 2000US-0185216P.

PR 12-MAY-2000; 2000US-0203785P.

PR 15-MAY-2000; 2000US-0204226P.

PR 16-MAY-2000; 2000US-0204525P.

PR 16-MAY-2000; 2000US-0204821P.

PR 16-MAY-2000; 2000US-0204908P.

PR 16-MAY-2000; 2000US-0205232P.

PR 17-MAY-2000; 2000US-0204815P.

PR 17-MAY-2000; 2000US-0204863P.

PR 17-MAY-2000; 2000US-0205221P.

PR 17-MAY-2000; 2000US-0205285P.

PR 17-MAY-2000; 2000US-0205286P.

PR 17-MAY-2000; 2000US-0205287P.

PR 17-MAY-2000; 2000US-0205323P.

PR 17-MAY-2000; 2000US-0205324P.

XX (INCY-) INCYTE GENOMICS INC.

XX Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;

PI Chen A, D'sa SA, Ameshey S, Dahl CR, Dam TC, Daniels SE, Dufour GE;

PI Flores V, Fong WT, Greenawalt LB, Hillman JL, Jones AL, Liu TF;

PI Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daffo A;
PI Wright RJ, Yap PE, Yu JY, Bradler DL, Bratcher SR, Chen W;
PI Cohen HU, Hodgson DM, Lincoln SE, Jackson S;
XX WPI: 2001-502867/55.
DR P-PSDB; AAU19431.
XX

PT Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
PT enzymes, hormones and receptors, useful in diagnostics and therapeutics.
XX
PS Claim 1; Page 304; 522pp; English.

XX The invention relates to polynucleotides (I) encoding diagnostic and
CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes, and
CC proteins involved in growth and development and receptors. (I) and (II)
CC may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate DITHP expression. For example, (I) and (II)
CC may be used to treat disorders associated with decreased polypeptide
CC expression by rectifying mutations or deletions in a patient's genome,
CC that affect the activity of the DITHPs, by expressing inactive proteins
CC or supplementing the patient's own production of them. (I) and (II) may
CC be used to treat diseases, for example, cell proliferative disorder,
CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,
CC (I) may be used to produce the DITHPs, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the protein. (I) and
CC its complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids in
CC samples, and therefore which patients may be in need of restorative
CC therapy. (II) may also be used as antigens in the production of
CC antibodies against DITHPs and in assays to identify modulators of DITHP
CC expression and activity. The anti-DITHP antibodies and antagonists may
CC also be used to down regulate expression and activity. The anti-DITHP
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbent assay
CC (ELISA)). AAS30986-AA331196 represent human diagnostic and therapeutic
CC (DITHP) polynucleotides of the invention

XX Sequence 1393 BP; 375 A; 352 C; 304 G; 360 T; 0 U; 2 Other;

Query Match 58.6%; Score 778.8; DB 4; Length 1393;
Best Local Similarity 87.2%; Pred. No. 1.1e-244;
Matches 904; Conservative 0; Mismatches 112; Indels 21; Gaps 4;

QY 302 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAACCCAGATGAGTCACTGACTAA 361
Db 1 GAAGCTGTAAAGCTACTAATTTCTTCAAAGAGAGCCAGATGAGTCACTGACTAA 60
QY 362 ATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGATGTTAATGACTGAA 421
Db 61 ATCTACGACAGACCCCTGGACCGGCTGTAGCCCATGACCCATGTTAATGATATGAN 120
QY 422 GTACCCCTCCGAGGAAATCTCACTGGACACACCCCTACTACATCCATTCAGTAGGA 481
Db 121 AGCACTCC-CCCAAGGAAATTTCACTGACACACCCCTACTACCCCAATTCAGTAGGA 179
QY 482 AGCAGTTAGAGCAGTGTGTGAGCAACCTTCCCAACAGTACTTGGGTTTCTCTGTGAGAG 541
Db 180 AGCAGTTAGAGCGGTGTGTGAGCAACCTTCCCAACAGTACTTGGGTTTCTCTGTGAGAG 239
QY 542 GGTGGACTGAGAGACAGGACTAGCTGGATTTCTTAGCTGACTAAGATCCCAAGCCTA 601
Db 240 CGGGGACTGAGAGACAGGACTAGCTGGATTTCTTAGCTGACTAAGATCCCAAGCCTA 299
QY 602 NCTGGGAGGTGACCGCATCCATCTTTAAACATGGGGCTTGGCACTTAGCTCACCCGA 661
Db 300 GCTGGGAAGGTGACCGCGTCTACCTTTAAACACGGGGCTTGGCACTTAGCTCACCCGA 359
QY 662 CCAATC-----AGAGAGCTCACTAAATCTAATCAGGCAAAACAGGAGGTAAAG 712
Db 360 CCATCAGGTAGTAAAGAGAGCTCACTAAATCTAATTAGGCAAAACAGGAGGTAGAG 419
QY 713 CAATAGCCAATCATCTATTGCTTGAGAGCAACAGGGGAAGGCAAGGATTTGGGATATAA 772

```
Db 420 AATAGCAATCATCTATCGCTGAGAGCAGCAGAGGAGCAATGATCCGATATAA 479
Qy 773 CTAGGCAATCAAGCCAGCAAGCAACCCCTTTGGTCCCTCCCATTTGATGGAGC 832
Db 480 CCAAGCAATCGAGCCAGCAAGCGCTGCCCCCTTTGGTCCCTCCCTTTGATGGAGC 539
Qy 833 TCTGTTTCACTCTATTTCATCTATTAAATCATGCACTGCACTCTTCTGGTCCGTGT 892
Db 540 TCTGT-----TTTCACTCTATTAAATCTTGCACACTGCACTCTTCTGGTCTGTGT 589
Qy 893 TTTTATGGCTCAAGCTGAGCTTTTGTTCGCGCATCCACCACTGCTGTTTGCACCGTCACA 952
Db 590 TGTACGTTTGGAGCTTTCGCTCGCGCTCCACCACTGCTGTTTGCAGCATCGGA 649
Qy 953 GACCGCTGCTGATTCATCCCTTTGGATCCAGCAGAGTGTCCACTGCTGCTGATCC 1012
Db 650 GACCTGCGCTGACTTCATCCCTCCGATCTGCGAGGTGTTCA-TGTCTCTGATCC 708
Qy 1013 AGCAGGTACCATTTGCCACTCCGATCAGGCTAAGGCTTGCCATTTCTGTCATGGC 1072
Db 709 AGAGAGGCACCCATTGCCATTCTCGATTGGGCTTAAAGGCTTGCCATTGTTCTGCANGAC 768
Qy 1073 TAAGTGCTCGGTTTGTCTTAATAGAACTCAACACTGCTCACTGGGTTCCATGTTCTCT 1132
Db 769 TAAGTGCCCGGTTTATCTTAATCGAGCTGAACACTAGTTCGCTGGGTTCCACGATTCTCT 828
Qy 1133 TCCATGACCCACGCTTCTTAATAGAGCTATAAAGCTCACCGCATGGCCCAAGATTCCATT 1192
Db 829 TCCGTGACCCACGACTTCTAATAGAGCTATAAAGCTCACCGCATGGCCCAAGATTCCATT 888
Qy 1193 CTTGTTGATCTGAGGCCAAGAACCCAGGTGAGAGANGTGAAGCTTGCCACCATTTG 1252
Db 889 CTTTGAATCCGTGAGGCCAAGAACCCAGGTGAGAGAACATGAGGCTTGCCACCATTTT 948
Qy 1253 GGAAGTGCCCATGTCATTTGTTAGCGGCCACCACTCTGGAGCTGTGGAGCA 1312
Db 949 GGAAGTGCCCTGCGGCATTTTGAAGTGCCCTGCCACCATCTTGGAGCTCTGGAGCA 1008
Qy 1313 AGGATCCCCCAGTAACA 1329
Db 1009 AGGACCCCTGTAACA 1025

RESULT 10
AAA63826
ID AAA63826 standard; DNA; 2030 BP.
XX AC AAA63826;
XX
XX 06-AUG-2003 (revised)
XX 04-DEC-2000 (first entry)
XX
DE Nucleotide sequence of the MSRV-1 3' env and LTR regions.
XX
XX MSRV-1; pol region; long terminal repeat; LTR; RUS region; retrovirus;
XX ss.
XX
XX Multiple sclerosis associated retrovirus.
XX
FH Key Location/Qualifiers
FT CDS 1..1629
FT FT /*tag= a
FT FT /note= "Contains one termination codon"
FT sig_peptide 1..81
FT CAAT_signal 1800..1807
FT FT /*tag= c
FT CAAT_signal 1858..1864
FT FT /*tag= d
FT TATA_signal 1906..1911
FT FT /*tag= e
FT polyA_signal 1996..2002
```

```
FT /*tag= f
XX WO200047745-A1.
PN 17-AUG-2000.
XX
XX 15-FEB-2000; 2000WO-IB000159.
XX
XX 15-FEB-1999; 99EP-00420041.
XX
XX (INMR ) BIO MERIEUX.
XX
XX Paranhos-Baccala G, Perron H, Komurian-Pradel P;
XX
XX WPI; 2000-506097/45.
XX
XX P-PSDB; AAB08195.
XX
XX Nucleotide fragment of LTR-RUS region from Multiple Sclerosis retrovirus
XX (MSRV) used to detect the presence of MSRV-1 retrovirus in a biological
XX sample.
XX
XX Disclosure; Fig 2; 23pp; English.
XX
XX The present sequence represents the nucleotide sequence corresponding to
XX the 3' env region and long terminal repeat sequences from clone CL6 of
XX Multiple Sclerosis retrovirus (MSRV-1). The specification describes a
XX long terminal repeat (LTR)-RUS region which encodes the expression of a
XX MSRV-1 protein. This is unusual for LTRs, in particular in the RUS
XX region. The sequence includes CAAT and TATA signals which are present in
XX the U3 and R regions and are not directed towards the CDS indicated in
XX the features table. Probes and antibodies to the MSRV-1 retrovirus
XX protein and encoding polynucleotide sequences are used to detect the
XX presence of MSRV-1 retrovirus in a biological sample. (Updated on 06-AUG-
XX 2003 to correct OS field.)
XX
XX Sequence 2030 BP; 574 A; 559 C; 387 G; 510 T; 0 U; 0 Other;
```

```
Query Match 58.3%; Score 775.4; DB 3; Length 2030;
Best Local Similarity 92.8%; Pred. No. 1.8e-243;
Matches 812; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCAAGAGGGGGGAACTGTTTATTTT 60
Db 1140 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCAAGAGGGGGGAACTGTTTATTTT 1199
Qy 61 AGGGGAAGAGTGTGTAGTATGTTAATCAATCGGAATCATTAATGAGAGGAGTTAAAGA 120
Db 1200 AGGAGAGAGAGCTGTTATTTATGTTAATCAATCGAGAGTTGCTGAGAGGAGTTAAAGA 1259
Qy 121 AATTTCAGATCGAATATAATGTAGACAGAGGACCTTCAAAAACACCTGCGGGGCT 180
Db 1260 AATTTCAGATCGAATATAATGTAGACAGAGGAGCTTCAAAAACACCGAGCTGGGGCT 1319
Qy 181 CCTCAGCAATGAGATGCGCTGAGACTCTCCCTTCTTAGGACCTCTAGCAGCTAATATT 240
Db 1320 CCTCAGCAATGAGATGCGCTGAGCTTCTCCCTTCTTAGGACCTCTAGCAGCTAATATT 1379
Qy 241 TTTTACTCTCTTTGGACCTGTTATCTTCAACTTCTTGTAAAGTTGTCTTCCAGAAAT 300
Db 1380 GTTACTCTCTTTGGACCTGTTATCTTAACTCTCTTGTAAAGTTGTCTTCCAGAAAT 1439
Qy 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAACCCAGATGCGATGACTAA 360
Db 1440 TGAAGCTGTAAAGCTACAGATGTTCTTCAAAATGGAACCCAGATGCGATGACTAA 1499
Qy 361 AATCTACCGTGAGACCCCTGAGCCGCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
Db 1500 GATCCACCGTGAGACCCCTGAGCCGCTGCTAGCCCATGCTCCGATGTTAATGACATTGA 1559
Qy 421 AGTCACCCCTCCGAGGAAATCTCAACTGCACACCCCTACTACACTCCCAATTCAGTAGG 480
Db 1560 AGGCACCCCTCCGAGGAAATCTCAACTGCACACCCCTACTATGCCCAATTCAGGGGG 1619
```

QY 481 AAGCAGTTAGAGCAGTTGTTCAGCCAACTCCCAACAGTACTTGGCTTTTCTCTGTGAGA 540
DB 1620 AAGCAGTTAGAGCGGTTCATCAGCCAACTCCCAACAGCAGTCTGGGTTTCTCTGTGAGA 1679
QY 541 GGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAAATCCCNAAAGCCT 600
DB 1680 GGGGGAGTGGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAAATCCCNAAAGCCT 1739
QY 601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTAGCTCACACCCCG 660
DB 1740 AGCTGGGAAGGTGACTGCATCCACCTCTAAACATGGGGCTTGCAACTAGCTCACACCCCG 1799
QY 661 ACCAATCAGAGACTCCTAAATGCTATCAGGCAAAACAGGAGGTAAAGCAATAGCC 720
DB 1800 ACCAATCAGAGACTCCTAAATGCTATCAGGCAAAATAGGAGGTAAAGCAATAGCC 1859
QY 721 AATCATCTATTGCTCAGAGACAGCGGGAAGGACAGGATTCGGGATATAAATCTCAGGCA 780
DB 1860 AATCATCTATTGCTCAGAGACAGCGGGAAGGACAGGATTCGGGATATAAATCTCAGGCA 1919
QY 781 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGTATGGGAGCTCTGTTTT 840
DB 1920 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGTATGGGAGCTCTGTTTT 1979
QY 841 CACTCTATTTCATCTATTAAATCATGCAACTGCA 875
DB 1980 CACTCTATTTCATCTATTAAATCTTGCAACTGAA 2014

RESULT 11

AD41225
ID AAD41225 standard; cDNA; 2074 BP.
XX
AC AAD41225;
XX
DT 30-OCT-2002 (first entry)
XX
DE Human EMBRY-2 cDNA.
XX
KW Human; embryogenesis associated protein; AIDS; reproductive disorder;
KW infertility; endometriosis; endometrial tumour; inflammatory disorder;
KW autoimmune disorder; acquired immune deficiency syndrome; transgenic;
KW ovarian tumour; contact dermatitis; placenta disorder; preeclampsia;
KW EMBRY-2; allergy; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CD CDS 64..1692
FT /*tag= a
FT /product= "EMBRY-2 protein"
FT sig_peptide 64..123
FT /*tag= b
FT mat_peptide 124..1689
FT /*tag= c
FT /product= "Mature EMBRY-2 protein"
XX
PN W0200248362-A2.
XX
XX 20-JUN-2002.
XX
XX 14-NOV-2001; 2001WO-US043956.
XX PF
XX 15-NOV-2000; 2000US-0249407P.
XX PR
XX (INCY-) INCYTE GENOMICS INC.
XX PA
XX Ramkumar J, Arvizu C;
XX WPI; 2002-537629/57.
XX DR P-PSDB; AAE25054.
XX
XX New polypeptides of human embryogenesis associated proteins for screening

PT modulators useful for treating or preventing disorders e.g.
PT endometriosis, infertility, allergy, preeclampsia.
XX
PS Claim 59; Page 96-97; 97pp; English.
XX
CC The invention relates to human embryogenesis associated proteins (EMBRY)
CC and nucleic acid molecules encoding such proteins. EMBRY sequences are
CC useful for screening modulators useful for treating or preventing
CC disorders associated with abnormal expression of EMBRY. The disorders
CC treated include reproductive disorders such as infertility,
CC endometriosis, endometrial or ovarian tumour, autoimmune/inflammatory
CC disorder such as acquired immune deficiency syndrome (AIDS), allergies,
CC contact dermatitis; disorders of the placenta such as preeclampsia, for
CC abruptio placentae etc. Sequences of the invention are also useful
CC analysing a proteome of a tissue or a cell type. EMBRY proteins are
CC useful as immunogens for preparing antibodies. Polynucleotides of the
CC invention are useful for creating knockin humanised animals or transgenic
CC animals to model human diseases. They are also used in gene therapy. The
CC present sequence is human EMBRY-2 cDNA
XX
SQ Sequence 2074 BP; 583 A; 567 C; 392 G; 531 T; 0 U; 1 Other;
Query Match 57.7%; Score 766.8; DB 6; Length 2074;
Best Local Similarity 93.0%; Pred. No. 1-2e-240;
Matches 812; Conservative 0; Mismatches 60; Indels 1; Gaps 1;
QY 1 TC AAAATCGAAGAGCTTTAGACTTGCTAACCGCAAAAGAGGGGAACTGTTTATTTT 60
DB 1203 TC AAAATCGAAGAGCTTTAGACTTGCTAACCGCAAAAGAGGGGAACTGTTTATTTT 1262
QY 61 AGGGGAAGAATGCTGTGTATGTATGTTTAAATCAATCTGGAATCATTACTGAGAAAGTTAAGA 120
DB 1263 AGGAGAAGAATGCTGTGTATGTATGTTTAAATCAATCTGGAATGTTGTTTAAAGTAAAGA 1322
QY 121 AATTCGAGATCGAATAATAATGTAGACGAGAGACCTTCAAAACACATGCACCTGGGGCCT 180
DB 1323 AATTCGAGATCGAATAATAATGTAGACGAGAGACCTTCAAAACACATGCACCTGGGGCCT 1382
QY 181 CCTCAGCAATGGATGCCCTGGGACTCTCCCTTTTAGGACCTCTAGCAGCTATAATTT 240
DB 1383 CCTCAGCAATGGATGCCCTGGGTTCTCCCTTTCTTAGGACCTCTAGCAGCTCTAATTT 1442
QY 241 TTTACTCTCTTTGGACCTGTATCTTCAACTCTCTTGTAGTTTGTCTCTTCCAGAAAT 300
DB 1443 GTTACTCTCTTTGGACCTGTATCTTAACTCTCTTGTAGTTTGTCTCTTCCAGAAAT 1502
QY 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAAACCCAGATGCAGTCCATGACTAA 360
DB 1503 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAAACCCAGATGCAGTCCATGACTAA 1562
QY 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGTATGTTTAAATGACATTTGA 420
DB 1563 AATCTACCGGACCCCTGGACCGGCTGTAGCCCATGCTCGATGTTTAAATGACATTTGA 1622
QY 421 AGTCACCCCTCCGAGGAAATCTCAACTGCAACACCCCTACTACCTCCATTCAGTAGG 480
DB 1623 AGGCACTCTCTCCAGGAAATCTCAACTGCAACACCCCTACTATGCCCCCAATTCAGCAGG 1682
QY 481 AAGCAGTTAGAGCAGTTGTTCAGCCAACTCCCAACAGTACTTGGGTTTCTCTGTGAGA 540
DB 1683 AAGCAGTTAGAGCGGTGCTGCTGAGTCAACCTCCCAACAGCAGCTTGGGTTTCTCTGTGAGA 1742
QY 541 GGGTGGACTGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAAATCCCNAAAGCCT 600
DB 1743 GGGGGAGTGGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAAATCCCNAAAGCCT 1802
QY 601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTAGCTCACACCCG 660
DB 1803 AGCTGGGAAGGTGACCGCGTCCACCTTTAAACATGGGGCTTGCAACTAGCTCACACCCA 1862
QY 661 ACCAATCAGAGACTCCTCAATAAATGCTAATCAGGCAAAACAGGAGGTAAAGCAATAGCC 720
DB 1863 ACCAATCAGAGACTCCTCAATAAATGCTAATCAGGCAAAACAGGAGGTAAAGCAATAGCC 1922

Db	1161	ATTAAATCTTTGCAACTGCACCTCTTCTGTGTCATGTTTGTACGGCTCGAGCTGAGCTTCT	1220
Qy	917	GTTGCGCCATCCACCACACTGCTGTTTGCCACCGTACAGACCCGCTGCTGACTTCCATCCCT	976
Db	1221	GCTTGAGTTCACCACTGCTGATGGTTGTATCGCAGACCCGGCGGCTGATGTCACCACT	1280
Qy	977	TTGGAATCCAGCAGAGTGTCCACTGTGCTCCTGATCCAGCAGGTAACCATTCGCACATCCC	1036
Db	1281	CCGGATCCGGCAGGGTGTCCACTGTGCTCCTGATCCAGAACGGCGCAAATTGCGCTCCT	1340
Qy	1037	GATCAGGCTAAAGGCTTGCCATTGTTCTCTGCATGGCTAAGTGCTGGGTTGTGCTTAATA	1096
Db	1341	GATCGGCTTAAAGCTTGCCATTGTTCTCTGCAGGCTAAGTGCTGGGTTGTGCTTAATT	1400
Qy	1097	GAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTT - CCATGACCCACACGCG - TTCTAAT	1154
Db	1401	GAGCTGACACTAGTCACTGGGTTCCATGGTTGCTTCCGCTGACCCACGAGCTTCTAAT	1460
Qy	1155	AGAGCTATAACACTCACCGCATGCCCCAAGATGCCATTCTCTTGATATCTGTGAGGCCAAG	1214
Db	1461	AGAGCTGAACACTGTCATG -- GCCTCAAGATTCCGTTCTCTTGAATCCGTGAGGCCAAG	1517
Qy	1215	AA - CCCAGGTCAGAGAAGTGAAGCTTGCCACATTTGGGAAGTGCCCACTGCCATTT	1273
Db	1518	AACCCCAGGTCAGAGAACACGAGGCTTGCCATCATCTTGAAGTGGCCCCACGACCATC	1577
Qy	1274	TGG 1276	
Db	1578	TTG 1580	

RESULT 13
 AAX77526
 ID AAX77526 standard; cDNA; 2946 BP.
 XX
 AC AAX77526;
 XX
 DT 10-AUG-1999 (first entry)
 XX
 DE Human secreted protein AJ172 2 cDNA.

This invention describes novel human secreted proteins encoded by polynucleotides isolated from human adult testes, adult brain, adult blood or adult placenta, or murine adult bone marrow or thymus cDNA libraries. The products of the invention are predicted to have biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals, although no supporting data is given. Suggested activities include nutritional activity, cytokine and cell proliferation/differentiation activity, immune stimulating (e.g. as vaccines) or suppressing activity, haematopoiesis regulating activity, tissue growth activity, actinin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cadherin/tumour invasion suppressor activity, and tumour inhibition activity. The polynucleotides are also stated to be useful for gene therapy

Sequence 2946 Bp; 858 A; 801 C; 570 G; 717 T; 0 U; 0 Other;
xx
80

Query Match	55.1%	Score 731.8	DB 2	Length 2946	
Best Local Similarity	90.7%	Pred. No. 5e-229			
Matches 794	Conservative	0	Mismatches 69	Indels 12	Gaps 1
Qy	1	TCAAAATCGAAGAGCTTTAGACTTGTCTAAACCGCAAAGAGGGGGAACCTGTTTATATTTT		60	
Db	2067	TCAAAATCGAAGAGCTTTAGACTTGTCTAAACCGCTGAAGAGGGGGAACCTGTTTATATTTT		2126	
Qy	61	AGGGGAAGAATGCTGTTAGTATCTTAATCAATCTGGAATCATTTACTGAGAAAGTTAAAGA		120	
Db	2127	AGGGGAAGAATGCTGTTATTTATGTTAATCAATCCGGAATCGTCACTCAGAAAGTTAAAGA		2186	
Qy	121	AAATTGAGATCGAATATATATGTAGACGAGGACCTTCAAAAACACGTGCACCCCTGGGGCCT		180	
Db	2187	AAATTGAGATTCGAATACAACGTAGACGAGGAGCTTCGAAAACACGTGAGCCCTGGGGCCT		2246	
Qy	181	CCTCAGCCAAATGATGCGCCCTGGACTCTCCGCCCTTCTTAGGACCTCTAGCAGCTATAATATT		240	
Db	2247	CCTCAGCCAAATGATGCCCTTGATTTCTCCCTTCTTAGGACCTCTAGCAGCTATAATATT		2306	
Qy	241	TTTACTCCTCTTTGGACCCCTGATATCTTCAACTTCCCTTGTAAAGTTTGTCTCTTCCAGAAT		300	
Db	2307	GCTACTCCTCTTTGGACCCCTGTATCTTTAACTTCCCTTGTAAAGTTTGTCTCTTCCAGAAT		2366	
Qy	301	TGAAGCTGTAAAGCTACAATAGTTCTTCAAAATGGAAACCCAGATGCAGTCCCATGACTAA		360	
Db	2367	CGAAGCTGTAAAGCTA-----CAAAATGGAGCCCAAGATGCAGTCCCAAGACTAA		2414	
Qy	361	AACTTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGTTTAAATGACATTTGA		420	
Db	2415	GATCTACCGCAGACCCCTGGACCGGCTGCTAGCCACGATCTGATGTTTAAATGACATCAA		2474	
Qy	421	AGTCACCCCTCCGAGGAATCTCAACTGCACAACCCCTACTACTCCTCAATTCAGTAGG		480	
Db	2475	AGGCACCCCTCCTTGAGGAATCTCAGCTGCACAACCTCTACTAGCGCCCAATTCAGCAGG		2534	
Qy	481	AAGCAGTTAGACAGTTGTCAGCCAACTCCCCAAACAGTACTTGGGTTTTTCTGTTTGAGA		540	
Db	2535	AAGCAGTTAGACGGTCTGCGCCAACTCCCCAAACAGCAGTCTAGGTTTTTCTGTTTGAGA		2594	
Qy	541	GGGTGGACTCAGAGACAGGACTAGCTGGATTTCTTAGGCTGATTAAGAAATCCCNAAAGCCT		600	
Db	2595	TGGGGGACTCAGAGACAGGACTAGCTGGATTTCTTAGGCTGACTTAAGAAATCCCTTAAGCCT		2654	
Qy	601	ANCTGGGAAGGTGACCGCATCTTTAAACATGGGCTTGGAACTTAGCTACACACCCG		660	
Db	2655	AGCTGGGAAGGTGACCACTTCCACCTTTAAACACGGGGCTTGGAACTTAGCTACACCTG		2714	
Qy	661	ACCAATCAGAGAGCTCACTAAAATGCTTAATCAGGCCAAAAACAGGAGGTTAAGCAATAGCC		720	
Db	2715	ACCAATCAGAGAGCTCACTAAAATGCTTAATAGCCAAAAACAGGAGGTTAAGCAATAGCC		2774	
Qy	721	AATCATCTATTGCTGAGACACAGCGGGGAAGACAAAGGATTGGGATATAAACTCAGGCA		780	
Db	2775	AATCATCTATTGCTGAGACACAGCAGGAGGACAATGATTCGGGATATAAAACCAAGTC		2834	

XX
PT New polynucleotides encoding secreted proteins.

PS Claim 13a; Page 100-101; 142pp; English.

XX | DB
2775 AATCACTCATTTGGCTTGAGAGCACAGCAGGAGGGGACAAATGATCGGGATATAATATACCCCAAGATC 2837

QY 781 TTCAAGCCAGCAGCAACCCCTTTGGTCCCTCCCATTTGATGGAGCTCTGTTTT 840
DB 2835 TTGAGCCGGAACGGAACCCCTTTGGTCCCTCCCTTTGATGGAGCTCTGTTTT 2894
QY 841 CACTCTATTTCACCTCTATTAAATCATGCAACTGCA 875
DB 2895 CATGCTATTTCACCTCTATTAACTTGCRACTGCA 2929

RESULT 14

ID AAZ59468
AAZ59468 standard; cDNA; 2946 BP.

XX AAZ59468;
DT 11-APR-2000 (first entry)

XX Human secreted protein AJ172_2 polynucleotide sequence.

XX Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;
KW placental pathology; metastasis inhibition; nutritional activity;
KW immune stimulator; haematopoiesis regulator; tissue growth;
KW tumour inhibitor; anti-inflammatory; clone AJ172_2; ATCC_98115;
KW gene therapy; ss.

XX Homo sapiens.

XX OS
XX PN WO960020-A1.

XX PD 25-NOV-1999.

XX PF 17-MAY-1999; 99WO-US010915.

XX PR 18-MAY-1998; 98US-00080478.

XX PR 20-OCT-1998; 98US-00175928.

XX PA (GEMY) GENETICS INST INC.

XX PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;

XX PI Merberg D, Mi S, Treacy M;

XX XX WPI; 2000-116311/10.

XX DR P-PSDB; AAY67313.

XX PS Claim 14; Page 107-108; 149pp; English.

XX This is the human secreted protein AJ172_2 nucleotide sequence, obtained from a human adult testes cDNA library. The invention relates to secreted human and murine proteins. The polynucleotides and proteins are predicted to have biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals. Detection of the levels of the proteins can be used for the diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents which modulate the expression or function of the proteins may be used for treating a neoplastic disease and inhibiting metastasis. Other suggested activities include nutritional activity (e.g. in feeds), cytokine and cell proliferation/differentiation activity, immune stimulating (e.g. as vaccines) or suppressing activity, haematopoiesis regulating activity, tissue growth activity, activin/inhibin activity.

CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity. The
CC polynucleotide sequences are also stated to be useful for gene therapy

SQ Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 U; 0 Other;

Query Match

Best Local Similarity 55.1%; Score 731.8; DB 3; Length 2946;

Matches 794; Conservative 0; Mismatches 69; Indels 12; Gaps 1;

QY 1 TCATAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAGAGGGGACCTGTTATTTTT 60
DB 2067 TCATAATCGAAGAGCTTTAGACTTGCTAACCGTGAAAGAGGGGAAACCTGTTATTTTT 2126
QY 61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTTACTGAGAAAGTTAAAGA 120
DB 2127 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCGTCACTGAGAAAGTTAAAGA 2186
QY 121 AATTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAAACATGCAACCTGGGGCCT 180
DB 2187 AATTGAGATCGAATATAACGTTAGAGCAGAGGAGCTTCGAAAACATGCAACCTGGGGCCT 2246
QY 181 CCTCAGCCAATGATGCCCTGGAGCTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATATT 240
DB 2247 CCTCAGCCAATGATGCCCTGGAGCTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATATT 2306
QY 241 TTTACTCCTCTTTGGACCCCTGATATCTTCAACTCTCTTTAAAGTTGCTCTCTCCCAAT 300
DB 2307 GCTACTCCTCTTTGGACCCCTGATATCTTAACTCTCTTTAACTTTGCTCTCTCCCAAT 2366
QY 301 TGAAGCTGTAAGCTACAAATAGTTCTTCAAAATGGAACCCAGATGCAATGCAATGCTAA 360
DB 2367 CGAAGCTGTAAGCTA-----CAAAATGGAACCCAGATGCAATGCAATGCTAA 2414
QY 361 AATCTACCGTGGACCCCTGGACCGGCTGTCTAGACTATGCTCTGATGTTAATGACATGA 420
DB 2415 GATCTACCGCAGACCCCTGGACCGGCTGTCTAGCCACGATCTGATGTTAATGACATCA 2474
QY 421 AGTCACCCCTCCGAGGAAATCTCACTGCACACCCCTACTACCTCCCAATTCAGTAGG 480
DB 2475 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACTCTACTACGCCCAATTCAGCAGG 2534
QY 481 AAGCAGTTAGAGCAGTTGTTCAGCCAACTCCCAACAGTACTTTGGGTTTTCTGTTGAGA 540
DB 2535 AAGCAGTTAGAGCGGTCGTCGGCCAACTCCCAACAGTACTTTGTTGTTGAGA 2594
QY 541 GGGTGAAGTGAAGCAGGACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCCNAAAGCCT 600
DB 2595 TGGGGAGCTGAGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCCTAAGCCT 2654
QY 601 ANCTGGGAAGTGAACCGCATCCATCTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
DB 2655 AGCTGGGAAGTGAACCATCCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2714
QY 661 ACCAATCAGAGAGCTCACTAAAAATGCTAATCAGGCAAAAAACAGAGGTAAGCAATAGCC 720
DB 2715 ACCAATCAGAGAGCTCACTAAAAATGCTAATCAGGCAAAAAACAGAGGTAAGCAATAGCC 2774
QY 721 AATCATCTATTGCTGAGAGCAGCGGGAAGCAAGGATTTGGGATATAAACTCAGGCA 780
DB 2775 AATCATCTATTGCTGAGAGCAGCGAGGAGGACAATGATCGGATATAAACCCCAAGTC 2834
QY 781 TTCAGCCAGCAACAGCAACCCCTTTGGTCCCTCCCAATTTGTATGGGAGCTCTGTTTT 840
DB 2835 TTCGAGCCGCAACCGCAACCCCTTTGGTCCCTCCCTTTGTATGGGAGCTCTGTTTT 2894
QY 841 CACTCTATTTCACCTCTATTAAATCATGCAACTGCA 875
DB 2895 CATGCTATTTCACCTCTATTAAATCTTGCACACTGCA 2929

RESULT 15

ADC38776

ID ADC38776 standard; cDNA; 2946 BP.

XX AC ADC38776;

XX DT 18-DEC-2003 (first entry)

XX DE Human cDNA encoding a secreted protein #63.

XX KW ss; gene; immune disorder; severe combined immunodeficiency; SCID;

XX AAD24195;
 AC 07-MAY-2002 (first entry)
 DT Human syncytin cDNA.
 DE Human; syncytin; preeclampsia; gestational trophoblast disorder;
 KW choriocarcinoma; hydatiform mole; placental site tumour; abortion;
 KW envelope gene; human endogenous defective retrovirus; HERV-W; ss.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FT CDS 930..2546
 FT /*tag= a
 FT /product= "Syncytin"
 XX WO200204678-A2.
 XX 17-JAN-2002.
 XX 09-JUL-2001; 2001WO-US021719.
 XX 07-JUL-2000; 2000US-0216657P.
 XX (GEMY) GENETICS INST INC.
 XX Keith JC, McCoy JM, Mi S;
 XX WPI; 2002-171727/22.
 XX P-PSDB; AAE14540.
 XX Identifying a compound for treating a subject with or at risk of
 XX developing preeclampsia, comprises determining whether the expression or
 XX activity of syncytin in the cell is modulated in the presence of a test
 XX compound.
 XX Disclosure; Page 39-42; 43pp; English.
 XX The invention relates to identifying compounds which are modulators of
 XX syncytin expression. The syncytin modulators are useful in diagnosis and
 XX treatment of preeclampsia and gestational trophoblast disorders (e.g.
 XX choriocarcinoma, hydatiform mole, placental site tumour and missed/
 XX incomplete abortion). Syncytin is a human gene derived from the envelope
 XX gene of human endogenous defective retrovirus, HERV-W. The present
 XX invention is based partly on the discovery that syncytin expression is
 XX dramatically reduced in preeclampsia, and is also mis-localised to the
 XX apical syncytiotrophoblast membrane. The present sequence is human
 XX syncytin cDNA
 XX Sequence 2930 BP; 842 A; 800 C; 571 G; 717 T; 0 U; 0 Other;
 XX
 XX Query Match 55.0%; Score 730.8; DB 6; Length 2930;
 XX Best Local Similarity 90.7%; Pred. No. 1.1e-228;
 XX Matches 793; Conservative 0; Mismatches 69; Indels 12; Gaps 1;
 XX
 XX 1 TCAAAATCGAAGAGCTTTAGACTTGTCAACCGCCAAAGAGGGGACCTGTTATTTT 60
 XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 2069 TCAAAATCGAAGAGCTTTAGACTTGTCAACCGCCAAAGAGGGGACCTGTTATTTT 2128
 XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 61 AGGGGAAGAATGCTGTATGTTATCAATCTGGAATCAATCTGAGAAAGTTAAAGA 120
 XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 2129 AGGGGAAGAATGCTGTATGTTATCAATCGGAATCGTCACTGAGAAAGTTAAAGA 2188
 XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 121 AATTGAGATCGAATAATATGTAGACAGAGGACCTTCAAAACATGACCTGGGCGCT 180
 XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 2189 AATTGAGATCGAATAACAGTAGACAGAGGAGCTTCGAAACATGACCTGGGCGCT 2248
 XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 181 CCTCAGCAATGATGCTGCTGCTCTCCCTTTAGGACCTCTAGCAGCTATAATTT 240
 XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX 2249 CCTCAGCAATGATGCTGCTGCTCTCCCTTTAGGACCTCTAGCAGCTATAATTT 2308
 XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 241 TTTACTCCTCTTTGGACCCCTGTATCTTCAACTTCTTGTAAAGTTTCTCTCTCCAGAT 300
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2309 GCTACTCCTCTTTGGACCCCTGTATCTTAAACCTCTTGTAACTTTCTCTCTCCAGAT 2368
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAAACCCAGATGACATGACTAA 360
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2369 CGAAGCTGTAAACTA-----CAATGGAGCCCAAGATGAGTCCAAGACTAA 2416
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGTCTGTATGTTAATGACTGA 420
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2417 GATCTACCGCAGACCCCTGGACCGGCTGTAGCCACGATCTGATGTTAATGACATCA 2476
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 421 AGTCACCCCTCCGAGGAATCTCACTGACACACCCCTCACTACACTCCATTCAGTAGG 480
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2477 AGGACCCCTCCGAGGAATCTCAGCTGCACACCTCTACTAGCCCCCAATTCAGCAGG 2536
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 481 AAGCAGTTAGAGCAGTTGTGAGCCAACTCCCAACAGTACTTGGGTTTCTCTCTGAGA 540
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2537 AAGCAGTTAGAGCGGCTCGTCGGCCNACCTCCCAACAGCAGCTAGGTTTCTCTGAGA 2596
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 541 GGTGAGCTGAGAGACAGGACTAGCTGGATTCTTAGCTGACTAAGAAATCCCNAGCCT 600
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2597 TGGGGACTGAGAGACAGGACTAGCTGGATTCTTAGCTGACTAAGAAATCCCTAAGCCT 2656
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 601 ANCTGGGAGGTGACCGCATCCATCTTAAACATGGGGCTTGCNACTTAGCTCACACCG 660
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2657 AGCTGGGAGGTGACCCACATCCACCTTTAAACAGGGGCTTGCNACTTAGCTCACACCTG 2716
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGCAAAAACAGAGGTAAGCAATAGCC 720
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2717 ACCAATCAGAGAGCTCACTAAATGCTAATCAGCAAAAACAGAGGTAAGCAATAGCC 2776
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 721 AATCATCTATTGCTGAGACACAGCGGGAAGGCAAGGATTGGGATATAAACTCAGGCA 780
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2777 AATCATCTATTGCTGAGACACAGCGGGAAGGCAAGGATTGGGATATAAACTCAGGCA 2836
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 781 TTCAAGCAGCAACAGCAACCCCTTTGGTCCCTCCCATTTGTATGGGAGCTCTGTTT 840
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2837 TTGAGCGGCAACAGCGCAACCCCTTTGGTCCCTCCCATTTGTATGGGAGCTCTGTTT 2896
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 841 CACTCTATTCTCACTCTATTAAATCATGCAACTGC 874
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2897 CATGCTATTCTCACTCTATTAAATCTTGCACACTGC 2930
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 XX RESULT 17
 XX AAF55630
 XX ID AAF55630 standard; DNA; 2781 BP.
 XX AC AAF55630;
 XX AC AAF55630;
 XX DT 29-MAY-2001 (first entry)
 XX DE Nucleotide sequence of a human endogenous retrovirus envelope protein.
 XX KW Envelope protein; HERV; syncytia formation; placental development;
 XX KW syncytia; cancer; cell adhesion; ss.
 XX OS Human endogenous retrovirus.
 XX FH Key Location/Qualifiers
 XX CDS 762..2378
 XX FT /*tag= a
 XX FT /product= "envelope protein"
 XX PN WO200116171-A1.
 XX 08-MAR-2001.
 XX 01-SEP-2000; 2000WO-FR002429.
 XX 01-SEP-1999; 99FR-00011141.
 XX 15-SEP-1999; 99FR-00011793.

XX (INMR) BIO MERIEUX.
PA (INRM) INST NAT SANTE & RECH MEDICALE.
XX Mallet F, Cosset F, Blond J, Lavillette D, Bouton O, Ruggieri A;
XX WPI; 2001-226676/23.
DR P-PSDB; AAB67652.
XX
XX Detecting expression of human endogenous retrovirus envelope protein in
PT cells of a tissue or culture, from its ability to induce syncytia.
XX
XX Disclosure; Page 44-45; 57pp; French.
XX
XX The present sequence encodes a human endogenous retrovirus envelope
CC protein. The specification describes a method for detecting expression of
CC an envelope protein from a human endogenous retrovirus (HERV), in cells,
CC of a tissue or culture. The method comprises detecting syncytia formation
CC due to the fusogenic properties of the envelope protein. Envelope
CC polypeptides and polynucleotides are used to produce therapeutic or
CC prophylactic compositions, particularly for treatment of cancer, to
CC correct defects in placental development for other natural formation of
CC other types of syncytia), and to promote adhesion of cells in grafts or
CC cellular repair processes. Expression of sequences antisense to the
CC polynucleotide are used to prevent formation of syncytia
XX
XX Sequence 2781 BP; 747 A; 770 C; 556 G; 708 T; 0 U; 0 Other;
XX
Query Match 54.9%; Score 730.2; DB 5; Length 2781;
Best Local Similarity 90.6%; Pred. No. 1.6e-228;
Matches 793; Conservative 0; Mismatches 70; Indels 12; Gaps 1;
1 TCAAAATCGAAGAGCTTTAGACTTGTCTTAAACCGGCAAGAGGGGGAACCTGTTTATTTT 60
1901 TCAAAATCGAAGAGCTTTAGACTTGTCTTAAACCGGCAAGAGGGGGAACCTGTTTATTTT 1960
61 AGGGGAAGAATGCTGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 120
1961 AGGGGAAGAATGCTGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2020
121 AATTGAGATCGAATACAGTGTAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 180
2021 AATTGAGATCGAATACAGTGTAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2080
181 CCTCAGCAATGATGCGCTGGAGCTCTCCCTCTTCTTAGGACCTCTAGCAGCTATATATT 240
2081 CCTCAGCAATGATGCGCTGGAGCTCTCCCTCTTCTTAGGACCTCTAGCAGCTATATATT 2140
241 TTTACTCTCTTTGGAGCCCTGATCTTCAACTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 300
2141 GCTACTCTCTTTGGAGCCCTGATCTTCAACTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2200
301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAACCCAGATGAGTCCATGACTAA 360
2201 CGAAGCTGTAAAGCTA-----CAAAATGGAACCCAGATGAGTCCATGACTAA 2248
361 AATTCTACCGTGGAGCCCTGGAGCCGCTGCTAGACTATGCTCTGATGTTAATGACATTTGA 420
2249 GATCTACCGAGAGCCCTGGAGCCGCTGCTAGCCACAGATCTGATGTTAATGACATCA 2308
421 AGTCACCCCTCCGAGGAATCTCAATGTCACAAACCCCTACTACATCTCAATTCAGTAGG 480
2309 AGGCACCCCTCCGAGGAATCTCAGTGTGACAACTCTACTACGCCCAATTCAGCAGG 2368
481 AAGCAGTTAGAGAGCTGTGAGCAACCTCCCAACAGTACTTGGGTTTCTTCTTCTTCTTCTTCT 540
2369 AAGCAGTTAGAGAGCTGTGAGCAACCTCCCAACAGTACTTGGGTTTCTTCTTCTTCTTCTTCT 2428
541 GGTGTGAGTACAGAGAGGAGTACTGATGTTTCTTAGGCTGACTAAGAATCCCNAGCCCT 600
2429 TGGGGAGTACAGAGAGGAGTACTGATGTTTCTTAGGCTGACTAAGAATCCCNAGCCCT 2488
601 ANCTGGGAAGGTGACCGCATCTTCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660

Db 2489 AGCTGGGAAGGTGACCAATCCACCTTTTAAACACAGGGGCTTGCAACTTAGCTCACACCTG 2548
QY 661 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAGGCAAGGAGGTAAAGCAATAGCC 720
Db 2549 ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAGGCAAGGAGGTAAAGCAATAGCC 2608
QY 721 AATCATCTATTGCTGAGAGCAGCGGGAAGGCAAGGATTGGGATATATAAATCAGGCA 780
Db 2609 AATCATCTATTGCTGAGAGCAGCGGGAAGGCAAGGATTGGGATATATAAATCAGGCA 2668
QY 781 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCTTATGTTGGAGCTCTGTTTT 840
Db 2669 TTGAGCCGCGCAACAGCAACCCCTTTGGGTCCCTCCCTTATGTTGGAGCTCTGTTTT 2728
QY 841 CACTCTATTCTCACTCTATTAAATCATGCAACTGCA 875
Db 2729 CATGCTATTCTCACTCTATTAAATCTTGCACACTGCA 2763
RESULT 18
AAS84210
ID AAS84210 standard; cDNA; 6394 BP.
XX AAS84210;
XX
DT 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #20014.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; as.
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; AAG20023.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1; SEQ ID NO 20014; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS4197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: the sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 6394 BP; 1840 A; 1597 C; 1384 G; 1571 T; 0 U; 2 Other;

Query Match 54.8%; Score 728.6; DB 5; Length 6394;

Best Local Similarity 78.6%; Pred. No. 9.1e-228;

Matches 1094; Conservative 0; Mismatches 223; Indels 74; Gaps 16;

```
QY 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAAACCGCAAAAGAGGGGGAACCTGTTTATTTT 60
DB 1175 TCAAAATCGAAGAGCTTTAGACTTGTCTAAACCGCTGAAGAGGGGGAACCTGTTTATTTT 1234

QY 61 AGGGGAAGATGCTGTTAGTATGTTAATCAATCTGGAATCATTAATCTGAGAAAGTTAAAGA 120
DB 1235 AGGGGAAGATGCTGTTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 1294

QY 121 AATTGAGATCGAATAAATGTAGAGCAGAGGACCTTCAAAACACTGCACCTGGGGCCT 180
DB 1295 AATTGAGATCGAATAAATGTAGAGCAGAGGAGCTTCGAAACACTGCACCTGGGGCCT 1354

QY 181 CCTCAGCAATGAGATGCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
DB 1355 CCTCAGCAATGAGATGCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1414

QY 241 TTTTACTCTCTTTGGACCCCTGTATCTTCACTTCTTGTAAAGTTTCTCTCTCCAGAAAT 300
DB 1415 GCTACTCTCTTTGGACCCCTGTATCTTCACTTCTTGTAAAGTTTCTCTCTCCAGAAAT 1474

QY 301 TGAAGCTGTAAGACTCAAAATAGTTCTTCAAAATGGAAACCCAGATGCAGTCCATGACTAA 360
DB 1475 CGAAGCTGTAAGACTCAAAATAGTTCTTCAAAATGGAAACCCAGATGCAGTCCATGACTAA 1522

QY 361 AATCTACCGTGGACCCCT-GGACGGCTGTAGACTATGCTCTGATGTTAAGTGAATG 419
DB 1523 GATCTACCGCAGACCCCTGGGACCGGCTGTAGTCCACAGTCTGATGTTAAGTGAATG 1582

QY 420 AAGTCACCCCTCCCGGGAATCTCAACTGCACACCCCTTACTACACTCC----- 469
DB 1583 AAGTCACCCCTCCCGGGAATCTCAACTGCACACCCCTTACTACACTCC----- 1642

QY 470 -----AATTCAGTAGGAAGCAGTTAGAGCAGTTGTCTAGCCAAACCTCCCAACAGTACTTG 524
DB 1643 CAGGGAAGCAGTTAGAGGCGGCTGTTGGGCGCAACCTTCCCAACAGCAGTACTTAGG 1702

QY 525 GGTTCCTCTGTTAGAGGGTGGACTGAGAG-----ACAGGACTAGCTGGA--TTTCCTAGG 578
DB 1703 GTTTTCTCTGTTGAGATGGGGGACTGAGGAGACAGGATAGCTGGGATTTCTCTTAGG 1762

QY 579 CTGACTAAGATCCCAAGCCTTACTGGGAAGTGCAG-----CATCCATCT 626
DB 1763 CTGACTAAGATCCCAAGCCTTACTGGGAAGTGCAG-----CATCCATCT 1822

QY 627 TTAACATGCGG-----CTTGCAACTTTAGCTCACACCC---GACCAATCAGAGAGCTCACTA 680
DB 1823 TAAACACAGGGGGCTTTGCACTTTTAGTTTCAACCTTGACCAATCAGAGAGCTCACTA 1882

QY 681 AAATG-CTAATCAGGCAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCTGA-G 738
DB 1883 AAATGCTAATTAGGCAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCTGAGG 1942

QY 739 AGCACAGGGAAGGACAGGATGGGATATAAATCACTCAGGATTCAGGCCAGC-AAACAGC 797
DB 1943 AGCACAGTGGGAGGACAGGATGGGATATAAATCACTCAGGATTCAGGCCAGGCAACGGC 2002

QY 798 AACCCCTTTGGGTCCCTCCCATTTGATGGAGCTCTGT-----TTTCACCTCTA 847
DB 2003 AACCCCTTTGGGTCCCTCCCATTTGATGGAGCTCTGT-----TTTCACCTCTA 2062
```

```
QY 848 TTTCACTCTATAAATCATCAACTG--CACTCTTCTGGTCCGCTGTTTTTTTATGCTCAA 905
DB 2063 TTTCACTCTATAAATCATCAACTG--CACTCTTCTGGTCCGCTGTTTTTTTATGCTCAA 2122

QY 906 GCTGAGCTTTGTTGCGCATCCACCACTGCTGTTGTCACCGTCA---CAGACCCGCTGC 962
DB 2123 GCTGAGCTTTGTTGCGCATCCACCACTGCTGTTGTCACCGTCA---CAGACCCGCTGC 2182

QY 963 TGACTTCCATCCCTTTGGATCCAGCAGAGTGTCACTGTGCTCTGATCCAGCGAGGTAC 1022
DB 2183 TGGCTTCCATCCCTTTGGATCCAGCAGAGTGTCTGTGTCATTCCTGATCCAGCGAGGCGC 2242

QY 1023 CCATTGCCACTCCCGATCAGGCTAAAGGCTTGCATTTGCTGCTGATGGCTTAAGTGCTG 1082
DB 2243 CCATTGCCACTCCCGATCAGGCTAAAGGCTTGCATTTGCTGCTGATGGCTTAAGTGCTG 2302

QY 1083 GGTTCCTTAATAGAACTGAAACACTGCTGCTCACTGGGTTCATGTTCTTCCAATGACCC 1142
DB 2303 GGTTCCTTAATAGAACTGAAACACTGCTGCTCACTGGGTTCATGTTCTTCCAATGACCC 2362

QY 1143 ACAGCTTCTAATAG--AGCTATTAACACTCACCGCATGGCCCAAGATTCATTCCTTGTA 1200
DB 2363 ATGCTTCTAATAGAACTAATTAACACTTACACATGGGCGCCCAAGATTCATTCCTTGTA 2422

QY 1201 --TCTGTGAGCCCAAGAACCCAGGTCAGAGAANGTGCAGCTTGCCACCATTTTGGGAGT 1258
DB 2423 ATTCCGTGAGGCCAAGAACCCAGGTCAGAGAANGTGCAGCTTGCCACCATTTTGGGAGT 2482

QY 1259 GGCCCACTGCTATTTGGTAGCGGCCCAACCACTCTGGGAGCTGTGGGAGCAAGGATC 1318
DB 2483 GGCCCGTGCATCTTTGGAAGCGGCTTGCCACCATTTTGGGAGCTGTGGGAGCAAGGATC 2542

QY 1319 CCCAGTAAACA 1329
DB 2543 CCCCGTAAACA 2553

RESULT 19
AAX25665
ID AAX25665 standard; cDNA to mRNA; 7582 BP.
XX
AC AAX25665;
XX
XX 21-MAY-1999 (first entry)
XX
DE Complete human endogenous retrovirus W genome.
XX
KW Clone; human endogenous retrovirus; genome; autoimmune disease;
KW multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KW disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
XX
OS Human endogenous retrovirus.
XX
PN WO9902696-A1.
XX
PD 21-JAN-1999.
XX
PF 06-JUL-1998; 98WO-FR001442.
XX
PR 07-JUL-1997; 97FR-00008815.
XX
PI (INMR ) BIO MERIEUX.
XX
PI Beseme F, Blond J, Bouton O, Mandrand B, Mallet F;
XX
XX WPI; 1999-120897/10.
XX
PT New nucleic acid sequences from human endogenous retrovirus-W - expressed
PT exclusively in placenta and useful in diagnosis and therapy of autoimmune
PT disease, and abnormal or failed pregnancy.
XX
XX Claim 1; Page 71-74; 106pp; French.
XX
```

CC This sequence represents the complete sequence of the human endogenous
CC retrovirus (HERV) W genome. The nucleic acids, their fragments or
CC peptides encoded by them are markers of autoimmune disease (e.g. multiple
CC sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,
CC insulin- dependent diabetes and related pathologies) and of abnormal or
CC unsuccessful pregnancy and can be used as chromosomal markers for
CC susceptibility to these conditions, or proximity markers of genes
CC associated with this susceptibility
XX
SQ Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 Other;

Query Match 54.4%; Score 722.6; DB 2; Length 7582;
Best Local Similarity 88.5%; Pred. No. 9.4e-226;
Matches 774; Conservative 19; Mismatches 70; Indels 12; Gaps 1;

QY 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAAGCGGCAAAAGAGGGGAACTGTTTATTTT 60
DB TCAAAATCGAAGAGCTTTAGACTTGTCTAAGCGGCAAAAGAGGGGAACTGTTTATTTT 6779

QY 61 AGGGGAAGAATGCTGTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
DB AGGGGAAGAATGCTGTATTTATTTAATCAATCTGGAATCGTCACTGAGAAAGTTAAAGA 6839

QY 121 AATTGAGTCGAATATAATGTAGACGAGAGGACCTTCAAAACACTGCACCTGGGCGCT 180
DB AATTGAGTCGAATATAATGTAGACGAGAGGACCTTCAAAACACTGCACCTGGGCGCT 6899

QY 181 CCTCAGCAATGAGTCCCTGGACTCTCCCTCTTAGGACCTCTAGCAGCTATAATATT 240
DB CCTCAGCAATGAGTCCCTGGACTCTCCCTCTTAGGACCTCTAGCAGCTATAATATT 6959

QY 241 TTTACTCCTCTTTGGACCCCTGTATCTTCAACTCTCTTGAAGTTTCTCTCCAGAAAT 300
DB TTTACTCCTCTTTGGACCCCTGTATCTTCAACTCTCTTGAAGTTTCTCTCCAGAAAT 7019

QY 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAATGGAAACCCAGATGAGTCCATGATAA 360
DB TGAAGCTGTAAAGCTACAAATAGTTCTTCAATGGAAACCCAGATGAGTCCATGATAA 7067

QY 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGATGTAATGACATTGA 420
DB AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGATGTAATGACATTGA 7127

QY 421 AGTCACCCCTCCGAGGAAATCTCAACTGCAACACCCCTACTACATCCCAATTCAGTAGG 480
DB AGTCACCCCTCCGAGGAAATCTCAACTGCAACACCCCTACTACATCCCAATTCAGTAGG 7187

QY 481 AAGCAGTTAGAGCAGTTGTGAGCCAACTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 540
DB AAGCAGTTAGAGCAGTTGTGAGCCAACTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 7247

QY 541 GGGTGGACTGAGACAGGACTAGCTGATTTCTAGCTGACTAAGAAATCCCAAGGCT 600
DB GGGTGGACTGAGACAGGACTAGCTGATTTCTAGCTGACTAAGAAATCCCAAGGCT 7307

QY 601 ANCTGGGAAGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTAGTCTACACCG 660
DB ANCTGGGAAGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTAGTCTACACCG 7367

QY 661 ACCAATCAGAGCTCATAAATGCTAATCAGGCMAAACAGAGGCTAAGCAATAGCC 720
DB ACCAATCAGAGCTCATAAATGCTAATCAGGCMAAACAGAGGCTAAGCAATAGCC 7427

QY 721 AATCATCTATTGCTGAGACAGCGGGAAGGCAAGGATTTGGATATATAACTCAGGCA 780
DB AATCATCTATTGCTGAGACAGCGGGAAGGCAAGGATTTGGATATATAACTCAGGCA 7487

QY 781 TTCAAGCAGCAACAGCAACCCCTTTGGGTCCCTCCCATCTGATGAGGAGCTCTGTTT 840
DB TTCAAGCAGCAACAGCAACCCCTTTGGGTCCCTCCCATCTGATGAGGAGCTCTGTTT 7547

QY 841 CACTCTATTTCACCTCTATTAAATCATGCAACTGCA 875
DB CACTCTATTTCACCTCTATTAAATCATGCAACTGCA 875

Db 7548 CATGCTATTCTACTCTATTAAATCTTGCACTGCR 7582

RESULT 20
AAAS9215
ID AAAS9215 standard; DNA; 7582 BP.
XX
AC AAAS9215;
XX
DT 07-NOV-2000 (first entry)
XX
DE Human endogenous retrovirus W (HERV-W) sequence.
XX
KW Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
KW gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
OS Human endogenous retrovirus.
XX
FH Key Location/Qualifiers
LTR 1. .120
FT /*tag= a
FT /note= "R of 5' LTR"
LTR 121. .575
FT /*tag= b
FT /note= "US of 5' LTR"
FT primer_bind 579. .596
FT /*tag= c
FT CDS 5581. .7194
FT /*tag= d
FT /note= "ORF1 env538"
FT CDS 7039. .7194
FT /*tag= e
FT /note= "ORF3 48 AA"
FT misc_feature 7244. .7254
FT /*tag= g
FT /note= "polypurine tract"
LTR 7256. .7582
FT /*tag= h
FT /note= "U3-R of 3' LTR"
FT polyA_signal 7583. .7569
FT /*tag= i
XX WO200043521-A2.
XX 27-JUL-2000.
XX
XX 21-JAN-2000; 2000WO-FR000144.
XX
XX 21-JAN-1999; 99FR-00000888.
XX (INMR) BIO MERIEUX.
XX
XX Paranhos-Baccala G, Mallet F, Voisset C;
XX
XX WPI; 2000-499229/44.
XX
XX New nucleic acid from human endogenous retrovirus, useful e.g. for
XX diagnosis of autoimmune disease and complications of pregnancy, contains
XX at least part of the gag gene.
XX
XX Disclosure; Page 49-52; 53pp; French.
XX
XX The present sequence represents an endogenous retrovirus, which is
XX associated with an autoimmune disease, and is integrated into the human
XX genome. The retrovirus is human endogenous retrovirus W (HERV-W). The
XX HERV-W retrovirus is associated with autoimmune disease, failure of
XX pregnancy or disorders of pregnancy. HERV-W nucleic acid fragments, or
XX proteins derived from it, are useful for diagnosis of autoimmune disease
XX (specifically multiple sclerosis) and for monitoring pregnancy. The
XX nucleic acid fragments may also be used for in situ labelling of isolated

CC chromosomes, while the transcription product can be used to study or
CC monitor T cell proliferation in vitro
XX
SQ Sequence 7582 BP; 2156 A; 1876 C; 1538 G; 1796 T; 0 U; 216 Other;
Query Match 54.4%; Score 722.6; DB 3; Length 7582;
Best Local Similarity 88.5%; Pred. No. 9.4e-226;
Matches 774; Conservative 19; Mismatches 70; Indels 12; Gaps 1;
Qy 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAAACCGCAAGAGGGGAACTGTTTATTTT 60
Db 6720 TCRAAATCGAAGAGCTTTAGACTTGTCTAAACCGCTGAGAGGGGAACTGTTTATTTT 6779
Qy 61 AGGGGAAGATGCTGTGTATGTTATCAATCTGGAATCAATTAATCAGAAAGTTAAAGA 120
Db 6780 AGGGGAAGATGCTGTGTATGTTATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 6839
Qy 121 AATTGTAGATCGAATATAATGTTAGACGAGGAGCTTCAAAACACCTGCACCTGGGCGCT 180
Db 6840 AATTCAGATCGAATACAAGTACGAGAGAGGCTTCGAAACACTGGACCTGGGCGCT 6899
Qy 181 CCTCAGCAATGAGTGGCCCTGAGCTCTCCCTTCTTAGGACCTCTAGCAGCTATAATTT 240
Db 6900 CCTCAGCCATGAGTGGCCCTGAGCTCTCCCTTCTTAGGACCTCTAGCAGCTATAATTT 6959
Qy 241 TTTACTCTCTTTGGACCTGTATCTTCAACTTCTTGTAAAGTTTGTCTCTCCAGAT 300
Db 6960 GCTACTCTCTTTGGACCTGTATCTTCAACTTCTTGTAAAGTTTGTCTCTCCAGAT 7019
Qy 301 TGAAGCTGTAAGCTACAATAGTCTTCAATGGAACCCAGATGAGTCCATGACTAA 360
Db 7020 CGAAGCTGTAAACTA-----CAAATGGAGCCCAAGATGCAAGTCCAAAGACTAA 7067
Qy 361 AATCTACGCTGGACCTGACCGCTGTAGACTATGCTCTCATGTTAATGACATTGA 420
Db 7068 GATCTACGAGACCTTGGACCGCTGTAGCCACGATCTGATGTTAATGACATCA 7127
Qy 421 AGTCACCCCTCCGAGGAAATCTCAACTGCAACACCCCTACTACACTCCAATTCAGTAGG 480
Db 7128 AGGCACCCCTCTGAGGAAATCTCAGCTGCAACACCTCTACTAGCCCCCAATTCAGCAGG 7187
Qy 481 AAGCAGTTAGAGCAGTGTGTCAGCAACCTCCCAACAGTACTTGGGTTTCTGTTGAGA 540
Db 7188 AAGCAGTTAGAGCAGTGTGTCAGCAACCTCCCAACAGTACTTGGGTTTCTGTTGAGA 7247
Qy 541 GGGTGGACTGAGACAGGACTAGCTGATTTCTAGCTGACTAAGAAATCCGNAAGCT 600
Db 7248 TGGGGGACTGAGACAGGACTAGCTGATTTCTAGCTGATTAAGAAATCCYTAAGCCT 7307
Qy 601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCG 660
Db 7308 AGSTGGGAGGTGACCATCCACCTTTAAACAGGGGCTTGCAACTTAGTTCACACCTG 7367
Qy 661 ACCAATCAGAGGCTCAATAATGCTAATCAGCGCAAAACAGAGGTAAGCAATAGCC 720
Db 7368 ACCAATCAGAGGCTCAATAATGCTAATTAGCGCAAAACAGAGGTAAGCAATAGCC 7427
Qy 721 AATCATCTTTGCTGAGACACAGCGGAGGACAGGATGGGATATAAATCACTCAGSCA 780
Db 7428 AATCATTTATTGCMGTGAGACACAGCGGAGGACAGGATGGGATATAAATCACTCAGSCA 7487
Qy 781 TTCAAGCAGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGATGGAGCTCTGTTTT 840
Db 7488 TTGAGCGGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGATGGAGCTCTGTTTT 7547
Qy 841 CACTCTATTTCACTCTATTAAATCATGCAACTGCA 875
Db 7548 CATGCTATTTCACTCTATTAAATCTTGCACTGCR 7582

RESULT 21
AAAX25660
ID AAX25660 standard; cDNA to mRNA; 1136 BP.

XX AX25660;
XX 21-MAY-1999 (first entry)
XX Human endogenous retrovirus W clone cl.C4C5.
XX
XX Clone; human endogenous retrovirus; genome; autoimmune disease;
XX multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
XX disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
XX Human endogenous retrovirus.
XX
XX W09902696-A1.
XX 21-JAN-1999.
XX 06-JUL-1998; 98WO-FR001442.
XX 07-JUL-1997; 97FR-00008815.
XX (INMR) BIO MERIEUX.
XX Beseme F, Blond J, Bouton O, Mandrand B, Mallet F;
XX WPI; 1999-120897/10.
XX New nucleic acid sequences from human endogenous retrovirus-W - expressed
XX exclusively in placenta and useful in diagnosis and therapy of autoimmune
XX disease, and abnormal or failed pregnancy.
XX Claim 1; Page 59-60; 106pp; French.
XX
XX This sequence represents clone cl.C4C5 of the human endogenous retrovirus
XX (HERV) W genome. The nucleic acids, their fragments or peptides encoded
XX by them are markers of autoimmune disease (e.g. multiple sclerosis,
XX rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
XX dependent diabetes and related pathologies) and of abnormal or
XX unsuccessful pregnancy and can be used as chromosomal markers for
XX susceptibility to these conditions, or proximity markers of genes
XX associated with this susceptibility
XX
XX Sequence 1136 BP; 336 A; 289 C; 241 G; 270 T; 0 U; 0 Other;
Query Match 54.3%; Score 721.2; DB 2; Length 1136;
Best Local Similarity 90.0%; Pred. No. 8.7e-226;
Matches 787; Conservative 0; Mismatches 75; Indels 12; Gaps 1;
Qy 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCAAGAGGGGAACTGTTTATTTT 60
Db 254 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCTGAAAGAGGGGAACTGTTTATTTT 313
Qy 61 AGGGGAAGATGCTGTGTATGTTAATCAATCTGGAATCAATTAATGAGAAAGTTAAAGA 120
Db 314 AGGGGAAGATGCTGTGTATGTTAATCAATCGGAATCGTCACTGAGAAAGTTAAAGA 373
Qy 121 AATTGTAGATCGAATATAATGTTAGACGAGGAGCTTCAAAACACCTGCACCTGGGCGCT 180
Db 374 AATTCAGATCGAATACAAGTACGAGCAGAGAGCTTCGAAACACTGGACCTGGGCGCT 433
Qy 181 CCTCAGCAATGAGTGGCCCTGAGCTCTCCCTTCTTAGGACCTCTAGCAGCTATAATTT 240
Db 434 CCTCAGCAATGAGTGGCCCTGAGCTCTCCCTTCTTAGGACCTCTAGCAGCTATAATTT 493
Qy 241 TTTACTCTCTTTGGACCTGTATCTTCAACTTCTTGTAAAGTTTGTCTCTCCAGAT 300
Db 494 GCTACTCTCTTTGGACCTGTATCTTAACTCTCTTGTAACTTTGTCTCTCCAGAT 553
Qy 301 TGAAGCTGTAAGCTACAATAGTCTTCAATCGAACCCAGATGAGTCCATGACTAA 360
Db 554 CGAAGCTGTAAACTA-----CAAATGGAGCCCAAGATGCAAGTCCAAAGACTAA 601
Qy 361 AATCTACGCTGGACCTGGACCGCTGTAGACTATGCTCTCATGTTAATGACATTGA 420

Db 602 GATCTACGAGACCCCTGGACCGGCTGCTAGCCACGATCTGATGTTAATGACATCA 661
Qy 421 AGTCACCCCTCCGAGGAAATCTCAACTGCACACACCCCTACTACACTCCCAATTCAGTAGG 480
Db 662 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACACCTCTACTAGCCGCCAATTCAGCAGG 721
Qy 481 AAGCAGTTAGAGCAGTTGTGAGCCAACTCCGCCAACAGTACTTGGGGTTTCTGTTGAGA 540
Db 722 AAGCAGTTAGAGCGGTGCTGGGCCAACCTCCCAACAGCAGCTTAGGTTTCTGTTGAGA 781
Qy 541 GGGTGGACTGAGACAGGAGTACTGATTTCTTAGCTGACTTAAGAATCCNAGACCT 600
Db 782 TGGGGGACTGAGACAGGAGTACTGATTTCTTAGCTGACTTAAGAATCCNAGACCT 841
Qy 601 ANCTGGGAAGGTGACCCGATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
Db 842 AGCTGGGAAGGTGACCATCCACCTTTAAACAGCGGGCTTGCAACTTAGCTCACACCTG 901
Qy 661 ACCAATCAGAGAGCTCTAAATGCTAATCAAGCAAAACAGAGGTAAAGCAATAGCC 720
Db 902 ACCAATCAGAGAGCTCACTAAATGCTAATTAGCAAAACAGAGGTAAAGCAATAGCC 961
Qy 721 AATCATCTATTGCTGAGACAGCGGAGGAGGAGGAGGATGGGATATAAATCAGGCA 780
Db 962 AATCATCTATTGCTGAGACAGCGGAGGAGGAGGAGGATGGGATATAAATCAGGCA 1021
Qy 781 TTCAAGCAGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGATGGGAGCTCTGTTT 840
Db 1022 TTGAGCGGAGCAACCGGCAACCCCTTTGGGTCCCTCCCATTTGATGGGAGCTCTGTTT 1081
Qy 841 CACTCTATTTCACTCTATTAAATCATCAACTGC 874
Db 1082 CATGCTATTTCACTCTATTAAATCTTCAGCTGC 1115

RESULT 22

AAAS59210
ID AAA59210 standard; DNA; 1136 BP.
AC AAAS59210;
XX
XX 07-NOV-2000 (first entry)
XX
XX 3' pol gene and 3' non coding sequences of HERV-W from human genome.
XX
XX Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
KW 989 gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
XX Homo sapiens.
XX
XX WO200043521-A2.
XX
XX 27-JUL-2000.
XX
XX 21-JAN-2000; 2000WO-FR000144.
XX
XX 21-JAN-1999; 99FR-00000888.
XX
XX (INMR) BIO MERIEUX.
XX
XX Paranhos-Baccala G, Mallet F, Voisset C;
XX
XX WPI; 2000-499229/44.
XX
XX New nucleic acid from human endogenous retrovirus, useful e.g. for
PT diagnosis of autoimmune disease and complications of pregnancy, contains
PT at least part of the gag gene.
XX
XX Disclosure; Page 46; 53pp; French.
XX
XX The present sequence represents an endogenous retroviral nucleic acid
CC fragment, which is associated with an autoimmune disease, and is

CC integrated into the human genome. The fragment is originally derived from
CC a novel retrovirus, human endogenous retrovirus W (HERV-W). The HERV-W
CC retrovirus is associated with autoimmune disease, failure of pregnancy or
CC disorders of pregnancy. The nucleic acid fragment, or proteins derived
CC from it, are useful for diagnosis of autoimmune disease (specifically
CC multiple sclerosis) and for monitoring pregnancy. The nucleic acid
CC fragments may also be used for in situ labelling of isolated chromosomes,
CC while the transcription product can be used to study or monitor T cell
XX proliferation in vitro
SQ Sequence 1136 BP; 336 A; 289 C; 241 G; 270 T; 0 U; 0 Other;
Query Match 54.3%; Score 721.2; DB 3; Length 1136;
Best Local Similarity 90.0%; Pred. No. 8.7e-226;
Matches 787; Conservative 0; Mismatches 75; Indels 12; Gaps 1;
Qy 1 TCAAAATCGAAGAGCTTTAGACTTGTCAACCGCAAAAGAGGGGAACTGTTTATTTT 60
Db 254 TCAAAATCGAAGAGCTTTAGACTTGTCAACCGCTGAAGAGGGGAACTGTTTATTTT 313
Qy 61 AGGGGAAGATGCTGTTAGTATTAATCAATCTGGAATCATCTGAGAAAGTTAAAGA 120
Db 314 AGGGGAAGATGCTGTTATTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 373
Qy 121 AATTTGAGATCGAATATAATGATAGACGAGAGCTTCAAAACACTGCACCTGGGCGCT 180
Db 374 AATTTGAGATCGAATATAACGATAGACGAGAGCTTCAAAACACTGCACCTGGGCGCT 433
Qy 181 CCTCAGCAATGATGATGCTGCTGCTCTCCCTCTTAGGACCTCTAGCAGCTATAATTT 240
Db 434 CCTCAGCAATGATGATGCTGCTGCTCTCCCTCTTAGGACCTCTAGCAGCTATAATTT 493
Qy 241 TTTACTCTCTTTGGACCCCTGTATCTTCAACTCTCTTGTGTTAAGTTGTCTCTCCAGAT 300
Db 494 GCTACTCTCTTTGGACCCCTGTATCTTAACTCTCTTGTGTTAAGTTGTCTCTCCAGAT 553
Qy 301 TGAAGCTGTAAGCTCAAAATAGTTCTTCAAAATGGAACCCAGATGAGTCCATGACTAA 360
Db 554 CGAAGCTGTAAGCTCAAAATAGTTCTTCAAAATGGAACCCAGATGAGTCCATGACTAA 601
Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTGATGTTAATGACATGA 420
Db 602 GATCTACCGCAGACCCCTGGACCGGCTGCTAGCCACGATCTGATGTTAATGACATCA 661
Qy 421 AGTCACCCCTCCGAGGAAATCTCAACTGCACAAACCCCTACTACACTCCCAATTCAGTAGG 480
Db 662 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACTCTACTAGCCGCCAATTCAGCAGG 721
Qy 481 AAGCAGTTAGAGCAGTTGTGAGCCAACTCCCAACAGTACTTGGGGTTTCTGTTGAGA 540
Db 722 AAGCAGTTAGAGCGGTGCTGGGCCAACCTCCCAACAGCAGCTTAGGTTTCTGTTGAGA 781
Qy 541 GGGTGGACTGAGACAGGAGTACTGATTTCTTAGCTGACTTAAGAATCCNAGACCT 600
Db 782 TGGGGGACTGAGACAGGAGTACTGATTTCTTAGCTGACTTAAGAATCCNAGACCT 841
Qy 601 ANCTGGGAAGGTGACCCGATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
Db 842 AGCTGGGAAGGTGACCATCCACCTTTAAACAGCGGGCTTGCAACTTAGCTCACACCTG 901
Qy 661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAACAGAGGTAAAGCAATAGCC 720
Db 902 ACCAATCAGAGAGCTCACTAAATGCTAATTAGCAAAACAGAGGTAAAGCAATAGCC 961
Qy 721 AATCATCTATTGCTGAGACAGCGGAGGAGGAGGAGGATGGGATATAAATCAGGCA 780
Db 962 AATCATCTATTGCTGAGACAGCGGAGGAGGAGGAGGATGGGATATAAATCAGGCA 1021
Qy 781 TTCAAGCAGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGATGGGAGCTCTGTTT 840
Db 1022 TTGAGCGGAGCAACCGGCAACCCCTTTGGGTCCCTCCCATTTGATGGGAGCTCTGTTT 1081
Qy 841 CACTCTATTTCACTCTATTAAATCATCAACTGC 874

```
Db 1082 CATGCTATTTCACCTCTATTAAATCTTCAGCTGC 1115
|||||
RESULT 23
AAAX25661
ID AAX25661 standard; cDNA to mRNA; 2782 BP.
XX
AC AAX25661;
XX
XX 21-MAY-1999 (first entry)
XX
XX Human endogenous retrovirus W clone cl.PH74.
XX
XX Clone; human endogenous retrovirus; genome; autoimmune disease;
XX multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
XX disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
XX
XX Human endogenous retrovirus.
XX
XX WO9902696-A1.
XX
XX 21-JAN-1999.
XX
XX 06-JUL-1998; 98WO-FR001442.
XX
XX 07-JUL-1997; 97FR-00008815.
XX
XX (INMR ) BIO MERIEUX.
XX
XX Beseme F, Blond J, Bouton O, Mandrand B, Mallet F;
XX WPI; 1999-120897/10.
XX
XX New nucleic acid sequences from human endogenous retrovirus-W - expressed
XX exclusively in placenta and useful in diagnosis and therapy of autoimmune
XX disease, and abnormal or failed pregnancy.
XX
XX Claim 1; Page 60-63; 106pp; French.
XX
XX This sequence represents clone cl.PH74 of the human endogenous retrovirus
XX (HERV) W genome. The nucleic acids, their fragments or peptides encoded
XX by them are markers of autoimmune disease (e.g. multiple sclerosis,
XX rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
XX dependent diabetes and related pathologies) and of abnormal or
XX unsuccessful pregnancy and can be used as chromosomal markers for
XX susceptibility to these conditions, or proximity markers of genes
XX associated with this susceptibility
XX
XX Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 U; 0 Other;
XX
Query Match 53.6%; Score 712.6; DB 2; Length 2782;
Best Local Similarity 89.4%; Pred. No. 1e-222;
Matches 782; Conservative 0; Mismatches 81; Indels 12; Gaps 1;
Qy 1 TC AAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAGAGGGGGAACTGTTTATTTT 60
Db 1902 TCGAATCGAAGAGCTTTAGACTTGCTAACCGCTGAGAGAGGGGGAACTGTTTATTTT 1961
Qy 61 AGGGGAAGATGCTGTTAGTATGTTATTAATCAATCTGGAATCAATTAAGTAAAGA 120
Db 1962 AGGGGAAGATGCTGTTATGTTATGTTATTAATCAATCGGAATCGTCACTGAGAAGTTGAAGA 2021
Qy 121 AATTTTGAGTCGATATAATGTTAGCAGAGGACCTTCAAAACACCTGACCTGGGCGCT 180
Db 2022 AATTCGAGATCGAATACACGATATAGCAGAGGAGCTTCGAAACACTGGACCTGGGCGCT 2081
Qy 181 CCTCAGCAATGAGTGCCTGACCTGCTCCCTCTTAGGACCTCTAGCAGCTATAATTT 240
Db 2082 CCTCAGCGGATGAGTGCCTGGATTTCCCTCTTAGGACCTCTAGCAGCTATAATTT 2141
Qy 241 TTTTACTCTCTTTGAGCCCTGATCTTCAACTCTCTGTTAAAGTTTGTCTTCCAGAA 300
|||||
```

```
2142 GCTACTCCTCTTTGGACCCCTGTATCTTTGACCTCCTTGTGTAACCTTCTCTTCCAGAA 2201
Qy 301 TGAAGCTGTAAAGCTACAATAATGTTCTTCAATGGAAACCCAGATGAGTCCATGACTAA 360
Db 2202 CGAAGCTGTGAAACTA-----CAATGGAGCCCAAGATGAGTCCCAAGACTAA 2249
Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTGTGATGTTAATGACATTTGA 420
Db 2250 GATCTACCGCAGACCCCTGGACCGGCTGTAGCCACGATCTGATGTTAATGACATCAA 2309
Qy 421 AGTCACCCCTCCGAGGAAATCTCAACTGCAACCCCTACTACACTCCCAATTCAGTAGG 480
Db 2310 AGGCACCCCTCCTGAGGAAATCTCAGCTGCAACCTCTACTAGCCCCCAATTCAGCAGG 2369
Qy 481 AAGCAGTTAGAGCAGTGTTCAGCAACCTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 540
Db 2370 AAGCAGTTAGAGCGGTGTCGGCAACCTCCCAACAGCACTTAGGTTTCTCTGTTGAGA 2459
Qy 541 GGGTGGACTGAGAGACAGGACTAGCTGGATTCTCTAGGCTGACTAAGAAATCCCNAGCCT 600
Db 2430 TGGGGACTGAGAGACAGGACTAGCTGGATTCTCTAGGCTGACTAAGAAATCCCTTAAGCCT 2489
Qy 601 ANCTGGGAAGGTGACCGCATCTTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
Db 2490 AGGTGGGAAGGTGACCATCTCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2549
Qy 661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAAAACAGGAGTAAAGCAATAGCC 720
Db 2550 ACCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAAAGACAGGAGTAAAGAAATAGCC 2609
Qy 721 AATCATCTATTGCTGAGACAGCGGAGGAGGAGGATGGGATATAAATCACTCAGGCA 780
Db 2610 AATCATTTATTGCTGAGACAGCAGGAGGAGGAGCAATGATCGGGATATAAATCACTCAGG 2659
Qy 781 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCTCCCAATTGTATGGGAGCTCTGTTT 840
Db 2670 TTGAGCGCGCAACGGCAACCCCTTTGGGTCCCTCCCTTTGTATGGGAGCTCTGTTT 2729
Qy 841 CACTCTATTTTCACTCTATTAAATCATGCAACTGCA 875
Db 2730 CATGCTATTTTCACTCTATTAAATCTTGCAACTGCA 2764
RESULT 24
AAAX59211
ID AAA59211 standard; DNA; 2782 BP.
XX
AC AAA59211;
XX
XX 07-NOV-2000 (first entry)
XX
XX 5' non coding, 3' pol gene, and 3' non coding sequences of HERV-W.
XX
XX Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
XX 9sg gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
XX Homo sapiens.
XX
XX WO200043521-A2.
XX
XX 27-JUL-2000.
XX
XX 21-JAN-2000; 2000WO-FR000144.
XX
XX 21-JAN-1999; 99FR-00000888.
XX
XX (INMR ) BIO MERIEUX.
XX
XX Paranhos-Baccala G, Mallet F, Voisset C;
XX WPI; 2000-499229/44.
XX
XX New nucleic acid from human endogenous retrovirus, useful e.g. for
```

PT diagnosis of autoimmune disease and complications of pregnancy, contains
PT at least part of the gag.gene.

XX Disclosure; Page 46-47; 53pp; French.

XX The present sequence represents an endogenous retroviral nucleic acid
CC fragment, which is associated with an autoimmune disease, and is
CC integrated into the human genome. The fragment is originally derived from
CC a novel retrovirus, human endogenous retrovirus W (HERV-W). The HERV-W
CC retrovirus is associated with autoimmune disease, failure of pregnancy or
CC disorders of pregnancy. The nucleic acid fragment, or proteins derived
CC from it, are useful for diagnosis of autoimmune disease (specifically
CC multiple sclerosis) and for monitoring pregnancy. The nucleic acid
CC fragments may also be used for in situ labelling of isolated chromosomes,
CC while the transcription product can be used to study or monitor T cell
CC proliferation in vitro

XX Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 U; 0 Other;

Query Match 53.6%; Score 712.6; DB 3; Length 2782;
Best Local Similarity 89.4%; Pred. No. 1e-222;
Matches 782; Conservative 0; Mismatches 81; Indels 12; Gaps 1;
Qy 1 TCAAAATCGAAGACTTTAGACTTGCTAACGCCAAAGAGGGGGAACCTGTTATTTT 60
Db 1902 TCGAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAGAGAGGGGGAACCTGTTATTTT 1961
Qy 61 AGGGGAAGATGCTTGTATGATGTTAATCAATCTGGAATCATTTACTGAGAAAGTTAAAGA 120
Db 1962 AGGGGAAGATGCTTGTATGATGTTAATCAATCTGGAATCATTTACTGAGAAAGTTGAAGA 2021
Qy 121 AATTGAGATCGAATAAATAGTAGAGCAGAGGACCTTCAAACACATCGCACCTGGGGCCT 180
Db 2022 AATTCCAGATCGAATAACAAGTATAGCAGAGGAGCTTCGAAACATCGACCTGGGGCCT 2081
Qy 181 CCTCAGCAATGAGTGGCTGGACTCTCCCTCTTTAGGACCTCTAGCAGCTATAATATT 240
Db 2082 CCTCAGCCGATGAGTGGCTGGATTCTCCCTCTTTAGGACCTCTAGCAGCTATAATATT 2141
Qy 241 TTTACTCTCTTTGGACCTGATCTTCACTTCTTGTAAAGTTCTCTCTCCCAAGAT 300
Db 2142 GCTACTCTCTTTGGACCTGATCTTCTGACCTCTTGTAACTTTGCTCTCCAGAAAT 2201
Qy 301 TGAAGCTGTAAAGCTCAAAATAGTTCTTCAAATGGAAACCCAGATGAGTCCATGACTAA 360
Db 2202 CGAAGCTGTGAACCTA-----CAAATGGAGCCCAAGATGAGTCCCAAGACTAA 2249
Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGTATGTTAATGACATGA 420
Db 2250 GATCTACCGCAGACCCCTGGACCGGCTGCTAGCCACGATCTGATGTTAATGACATCAA 2309
Qy 421 AGTCACCCCTCCGAGGAATCTCAACTGCACACCCCTACTACATCCCAATTCAGTAGG 480
Db 2310 AGGCACCCCTCCGAGGAATCTCAGCTGCACACCCCTACTAGCCCAATTCAGTAGG 2369
Qy 481 AAGCAGTTAGAGCAGTTGTGAGCCAACTCCCAACAGTACTTTGGGTTTTCCTGTTGAGA 540
Db 2370 AAGCAGTTAGAGCAGTTGTGAGCCAACTCCCAACAGCAGTCTAGGTTTTCCTGTTGAGA 2429
Qy 541 GGGTGGACTGAGAGCAGGACTAGCTGGATTCTTAGGCTGACTAAGAAATCCCNAGCCT 600
Db 2430 TGGGGGACTGAGAGCAGGACTAGCTGGATTCTTAGGCTGACTAAGAAATCCCTTAAGCCT 2489
Qy 601 ANCTGGGAAGGTGACCCGATCCATCTTTAAACATGGGGCTTGCACCTAGCTACACCCG 660
Db 2490 AGGTGGGAAGGTGACCATCCACCTTTAAACACGGGGCTTGCACCTAGCTACACCTG 2549
Qy 661 ACCAATCAGAGAGCTCACTAAATGCTAAATCAGGCAAAAACAGAGGCTAAAGCAATAGCC 720
Db 2550 ACCAATCAGAGAGCTCACTAAATGCTAAATGAGCAAGACAGGAGTAAAGAAATAGCC 2609
Qy 721 AATCATCTATTGCTGAGACACAGCGGAGGCAAGAGGATTTGGGATATAAACTCAGGCA 780
|||||

Db 2610 AATCATTTATTGCTGAGACACAGCAGGAGGACATGATCGGATATAAACCAAGTT 2669
Qy 781 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGATGGAGCTCTGTTT 840
Db 2670 TTGAGCGGCAACAGCGCAACCCCTTTGGGTCCCTCCCATTTGATGGAGCTCTGTTT 2729
Qy 841 CACTCTATTTCACCTCTATTAAATCATGCAACTGCA 875
Db 2730 CATGCTATTTCACCTCTATTAAATCTTGCAACTGCA 2764
RESULT 25
AAH20069
ID AAH20069 standard; DNA; 2782 BP.
XX
AC AAH20069;
XX
DT 08-AUG-2001 (first entry)
XX
DE HERV-W envelope protein G encoding nucleic acid.
XX
KW Human endogenous retrovirus; HERV-W; HERV; chromosome 7; env protein;
KW envelope protein; multiple sclerosis-related superantigen; vaccine;
KW surface antigen; transmembrane; multiple sclerosis; neuroprotective;
KW antisense-therapy; autoimmune disorder; ds.
XX
OS Human endogenous retrovirus.
XX
PH Key Location/Qualifiers
FT 5'UTR 1..762
FT CDS /*tag= a
FT 763..2379 /*tag= b
FT /product= "HERV-W envelope protein G"
FT /transl_except= (pos:790..792,aa:Phe)
FT /transl_except= (pos:793..795,aa:Thr)
FT /transl_except= (pos:812..814,aa:Leu)
FT /transl_except= (pos:818..820,aa:Ser)
FT /transl_except= (pos:862..864,aa:Tyr)
FT /transl_except= (pos:865..867,aa:Gln)
FT /transl_except= (pos:1174..1176,aa:Arg)
FT /transl_except= (pos:1441..1443,aa:Leu)
FT /transl_except= (pos:1903..1905,aa:Gln)
FT /transl_except= (pos:2017..2019,aa:Lys)
FT /transl_except= (pos:2026..2028,aa:Arg)
FT /transl_except= (pos:2044..2046,aa:Arg)
FT /transl_except= (pos:2089..2091,aa:Gln)
FT /transl_except= (pos:2170..2172,aa:Asn)
FT 3'UTR 2380..2782
FT /*tag= c
XX
WO200131021-A1.
XX
03-MAY-2001.
XX
30-OCT-2000; 2000WO-EP010659.
XX
28-OCT-1999; 99EP-00402690.
XX
(UYGE-) UNIV GENEVE.
XX
Conrad B, Mach B;
XX
WPI; 2001-316336/33.
DR P-PSDB; AAB75138.
XX
New human retrovirus HERV-W ENV proteins/peptides having superantigen
PT activity useful for diagnosing and treating multiple sclerosis.
XX
Claim 13; Fig 9; 94pp; English.
XX
On the basis of the PBS t-RNA motif used for the classification of human
CC endogenous retrovirus (HERVs) the full length endogenous provirus which

CC was been located on the long arm of human chromosome 7 (7q21-22) has been
CC designated HERV-W. The present invention describes proteins or peptides
CC (I) having superantigen (SAG) activity comprising the ENV protein (ENV)
CC of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)
CC have neuroprotective activity, and can be used in: vaccines; antisense-
CC therapy; and HERV-W SAG activity-inhibitors. (I) and encoding DNA/RNA are
CC useful for diagnosing multiple sclerosis (MS) or HERV-W-associated
CC disorders. (I) are also useful for identifying substances (and optionally
CC recovering) capable of binding to a retroviral superantigen associated
CC with MS, substances capable of blocking SAG activity and substances
CC capable of blocking transcription or translation of HERV-W retroviral
CC superantigen. A protein or peptide derived from (I), modified to be
CC devoid of SAG activity and being capable of generating an immune response
CC against HERV-W retroviral SAG is useful in therapy. Nucleic acid
CC molecules encoding (I) are useful as vaccines against MS. Substances
CC capable of blocking SAG activity, capable of binding to a retroviral
CC superantigen associated with MS, or capable of blocking transcription or
CC translation of HERV-W retroviral superantigen for use in treating or
CC preventing MS, obtained using (I) are useful for the treatment and
CC prevention of MS. (I) and nucleic acids encoding them are useful for
CC diagnosing autoimmune disease. The present sequence encodes the
CC specifically claimed envelope protein of HERV-W designated G
XX
SQ Sequence 2782 BP; 741 A; 768 C; 564 G; 709 T; 0 U; 0 Other;

Query Match 53.6%; Score 712.6; DB 5; Length 2782;

Best Local Similarity 89.4%; Pred. No. 1e-222;

Matches 782; Conservative 0; Mismatches 81; Indels 12; Gaps 1;

QY 1 TCRAAATCGAAGAGCTTTAGACTTGTCAACCGCCAAAGAGGGGACCTGTTATTTTT 60
DB 1902 TCRAAATCGAAGAGCTTTAGACTTGTCAACCGCTGAGAGAGGGGACCTGTTATTTTT 1961
QY 61 AGGGGAAGAATGCTGTATGTTATGTTATCAATCTGGAATCAATCTAGAGAAAGTTAAAGA 120
DB 1962 AGGGGAAGAATGCTGTATGTTATGTTATCAATCGGAATCGTCACTGAGAAGTTGAAGA 2021
QY 121 AATTGAGATCGAATAATAATGTTAGAGCAGAGAGACCTTCAAAAACACTGCACCTGGGSCCT 180
DB 2022 AATTCCAGATCGAATAACAAGTATAGCAGAGAGAGCTTCGAAAACACTGCAGCCTGGGSCCT 2081
QY 181 CCTCAGCAATGATGCGCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATATATT 240
DB 2082 CCTCAGCGATGATGCGCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATATATT 2141
QY 241 TTTACTCTCTTTGGACCTGTATCTTCAACTCTCTTTGTTAAAGTTTCTCTTCCAGAAAT 300
DB 2142 GCTACTCTCTTTGGACCTGTATCTTTGACCTCTTGTAACTTTGTCTCTTCCAGAAAT 2201
QY 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAAACCCAGATGCAGTCCATGACTAA 360
DB 2202 CGAAGCTGTGAACCTA-----CAAAATGGAGCCCAAGATGCAGTCCAAAGACTAA 2249
QY 361 AATCTACCTGGAGCCCTTGACCGGCTGTAGACTATGCTGATGTTAATGACATTGA 420
DB 2250 GATCTACCGCAGACCCCTTGACCGGCTGTAGCCACGATCTGATGTTAATGACATCAA 2309
QY 421 AGTCACCCCTCCGAGGAATCTCAACTGCAACACCCCTTACTACACTCCCAATTCAGTAGG 480
DB 2310 AGGCACCCCTCTTGAGGAATCTCAGCTGCAACACCTTCTACTAGCCGCCAATTCAGCAGG 2369
QY 481 AAGCAGTTAGAGCAGTTGTCAGCAACCTCCCCAACAGTACTTGGGTTTTCTGTTGAGA 540
DB 2370 AAGCAGTTAGAGCGGTGTCGCGCAACCTCCCCAACAGCAGTACTAGGTTTTCTGTTGAGA 2429
QY 541 GGGTGGACTGAGACAGGACTAGCTGATTTCTAGCTGACTAAGAAATCCCNAGACCT 600
DB 2430 TGGGGGACTGAGACAGGACTAGCTGATTTCTAGCTGACTAAGAAATCCCTTAAGCCT 2489
QY 601 ANCTGGGAAGTGACCGCATCTTTTAAACATGGGGCTTGCAACTTAGCTACACCCG 660
DB 2490 AGTGGGAAGTGACCGCATCTTTTAAACATGGGGCTTGCAACTTAGCTACACCTG 2549

QY 661 ACCAATCAGAGAGCTCACTAAATGCTTAATCAGCCAAACAGGAGCTAAAGCAATAGCC 720
DB 2550 ACCAATCAGAGAGCTCACTAAATGCTTAATCAGCCAAACAGGAGCTAAAGCAATAGCC 2609
QY 721 AATCATCTATTGCTCTGAGAGCAGAGCGGGAAGGACAGGATTGGGATATAAATCACTCAGGCA 780
DB 2610 AATCATCTATTGCTCTGAGAGCAGAGCGGGAAGGACAGGATTGGGATATAAATCACTCAGGCA 2669
QY 781 TTCAAGCCAGCAACAGCAACCCCTTTGGTTCCTCCCTTGTATGGAGCTCTGTTTT 840
DB 2670 TTCAAGCCAGCAACAGCAACCCCTTTGGTTCCTCCCTTGTATGGAGCTCTGTTTT 2729
QY 841 CACTCTATTTCCTCTCTATTAAATCAATCACTCAACTGCA 875
DB 2730 CATGCTATTTCCTCTCTATTAAATCTTGCACACTGCA 2764

RESULT 26

ID AAI14608 standard; DNA; 1894 BP.

XX AC AAI14608;

XX DT 12-OCT-2001 (first entry)

XX DE Probe #4541 for gene expression analysis in human cervical cell sample.

XX KW Probe; human; microarray; gene expression; cervical epithelial cell;

XX OS Homo sapiens.

XX PN WO200157278-A2.

XX PD 09-AUG-2001.

XX PP 30-JAN-2001; 2001WO-US000670.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488901/53.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human cervical epithelial cells.

XX PS Claim 25; SEQ ID NO 4541; 487pp; English.

XX CC The present invention relates to human single exon nucleic acid probes (SENPs). The present sequence is one such probe. The SENPs are derived from human HeLa cells. The SENPs can be used to produce a single exon microarray, which can be used for measuring human gene expression in a sample derived from human cervical epithelial cells. By measuring gene expression, the probes are therefore useful in grading and/or staging of diseases of the cervix, notably cervical cancer. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences)

XX SQ Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;

Query Match 53.0%; Score 704.8; DB 4; Length 1894;

Best Local Similarity 89.7%; Pred. No. 3e-220;

Matches 770; Conservative 0; Mismatches 79; Indels 9; Gaps 1;

```
Qy 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCCAAAGAGGGGAACTGTTTATTTT 60
Db 961 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCGGAAGAGGGGAACTTTTATTTT 1020
Qy 61 AGGGGAAGAAATGCTGTTAGTATGTTAATCAATCTGGAATCATTAATGAGAAAGTTAAAGA 120
Db 1021 AGAGGAAATGCTGTTGTTATGTTAATCAATCGGAATCATCACCGAAGAAAGTTAAAGA 1080
Qy 121 AATTGAGATCGAATATAATGTTAGAGCAGAGGACCTTCAAAACACTGCACCTCGGGCCT 180
Db 1081 AATTCAAGGTGCAATATAACCTAGAGCAAGAGGAGCTGCNAAAACACTGCACCTCGGGCCT 1140
Qy 181 CCTCAGCAATGGAATGCGCTGGAATCTCCCTCTTCTAGGACCTCTAGCAGCTATAATTT 240
Db 1141 CCTCAGCAATGGAATGCGCTGGAATCTCCCTCTTCTAGGACCTCTAGCAGCTATAATTT 1200
Qy 241 TTATCTCTCTTTGGACCTGTATCTTCAACTCTCTTTGTTAAAGTTGCTCTTCCAGAA 300
Db 1201 GTTACTCTCTTTGGACCTGTATCTTCAACTCTCTTTGTTAAAGTTGCTCTTCCAGAA 1260
Qy 301 TGAAGCTGTAAGCTACAAATAGTTCTTCAATGGAACCCAGATGAGTCCATGACTAA 360
Db 1261 CGAAGCAGTAAACTACAAATCGTTCTTCAATGGAACCCAGATGAGTCCATGACTAA 1320
Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
Db 1321 AATCTACCGGACCCCTGGACCGGCTGCTAGCCATGCTCTGATGTTAATGACATCA 1380
Qy 421 AGTCACCCCTCCGAGGAAATCTCAACTGACAAACCCCTACTACACTCCAAATCAGTAGG 480
Db 1381 AGGCACCCCTCCGAGGAAATCTCAACTGACAAACCCCTACTACACTCCAAATCAGTAGG 1440
Qy 481 AAGCAGTTAGACGATGTTGACGCAACCTCCGACAGTACTGGGTTTCTGTTGAGA 540
Db 1441 AAGCAGTTAGACGATGTTGTTGGCCAACTCCGACAGTACTGGGTTTCTGTTGAGA 1500
Qy 541 GGGTGGACTGAGACAGGACTAGCTGGATTTCTAGGCTGACTAAGAAATCCNAGCCT 600
Db 1501 GGGGGGACTGAGACAGGAAATAGTATTTCTAGCACTAAGAAATCCCTAAGACT 1560
Qy 601 ANCTGGAAGTGACCGCATCCATCTTTAAACATGGGGCTTGGAACTTAGCTCACACCG 660
Db 1561 AGCTGGAAGTGACCGCTTCCACCTTTAAACACCGGGCTTGCAACTTAGCTCACGCGCA 1620
Qy 661 ACCAATC-----AGAGAGCTCACTAATGCTAATCAGGCAAAACAGGAGTAA 711
Db 1621 ACCAATCAGATCTAAAGAGAGCTCACTAATGCTAATAGGCAAAACAGGAGTAA 1680
Qy 712 GCAATAGCAATCATCTATTGCTGAGACAGCGGGAAGGACAAGGATTTGGGATATA 771
Db 1681 GAATAGCAATCATCTATTGCTGAGACAGCGGGAAGGACAATGATCGGATATA 1740
Qy 772 ACTCAGGATTCAGCCAGCAACAGCAACCCCTTTGGGTCCTCCCATTTGATGGAG 831
Db 1741 ACCCAGGATTCGAGCAGCTACAGTACCCTCTTTGGGTCCTCCCTTTGATGGAG 1800
Qy 832 CTCTGTTTCACTCTATT 849
Db 1801 CTCTGTTTCACTCTATT 1818
```

RESULT 27
ABA56337
ID ABA56337 standard; DNA; 1894 BP.
XX
AC ABA56337;
XX
DT 01-FEB-2002 (first entry)
XX Human foetal liver single exon nucleic acid probe #4642.
DE Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
KW

```
XX Homo sapiens.  
OS WC2000157277-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX 30-JAN-2001; 2001WO-US000669.  
PF  
XX 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX (MOLB-) MOLECULAR DYNAMICS INC.  
PA Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483447/52.  
PI  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human foetal liver.  
PS  
XX Claim 1; SEQ ID NO 4642; 639pp + Sequence Listing; English.  
CC The invention relates to a single exon nucleic acid probe for measuring  
CC human gene expression in a sample derived from human foetal liver. The  
CC single exon nucleic acid probes may be used for predicting, measuring and  
CC displaying gene expression in samples derived from human foetal liver. The  
CC present sequence is a single exon nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published_pct_sequences  
XX  
SQ Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;  
Query Match 53.0%; Score 704.8; DB 4; Length 1894;  
Best Local Similarity 89.7%; Pred. No. 3e-220;  
Matches 770; Conservative 0; Mismatches 79; Indels 9; Gaps 1;  
Qy 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCCAAAGAGGGGAACTGTTTATTTT 60  
Db 961 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCGGAAGAGGGGAACTTTTATTTT 1020  
Qy 61 AGGGGAAGAAATGCTGTTAGTATGTTAATCAATCTGGAATCATTAATGAGAAAGTTAAAGA 120  
Db 1021 AGAGGAAATGCTGTTGTTATGTTAATCAATCGGAATCATCACCGAAGAAAGTTAAAGA 1080  
Qy 121 AATTGAGATCGAATATAATGTTAGAGCAGAGGACCTTCAAAACACTGCACCTCGGGCCT 180  
Db 1081 AATTCAAGGTGCAATATAACCTAGAGCAAGAGGAGCTGCNAAAACACTGCACCTCGGGCCT 1140  
Qy 181 CCTCAGCAATGGAATGCGCTGGAATCTCCCTCTTCTAGGACCTCTAGCAGCTATAATTT 240  
Db 1141 CCTCAGCAATGGAATGCGCTGGAATCTCCCTCTTCTAGGACCTCTAGCAGCTATAATTT 1200  
Qy 241 TTATCTCTCTTTGGACCTGTATCTTCAACTCTCTTTGTTAAAGTTGCTCTTCCAGAA 300  
Db 1201 GTTACTCTCTTTGGACCTGTATCTTCAACTCTCTTTGTTAAAGTTGCTCTTCCAGAA 1260  
Qy 301 TGAAGCTGTAAGCTACAAATAGTTCTTCAATGGAACCCAGATGAGTCCATGACTAA 360  
Db 1261 CGAAGCAGTAAACTACAAATCGTTCTTCAATGGAACCCAGATGAGTCCATGACTAA 1320  
Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420  
Db 1321 AATCTACCGGACCCCTGGACCGGCTGCTAGCCATGCTCTGATGTTAATGACATCA 1380  
Qy 421 AGTCACCCCTCCGAGGAAATCTCAACTGACAAACCCCTACTACACTCCAAATCAGTAGG 480
```

Db 1381 AGGCACCCCTCCGAGGAAATCTCAACTGACAAACCTTACTACGCCCAATTCAGCAGG 1440
Qy 481 AAGCAGTTAGAGCAGTTGTTCAGCCAACTCCGCCAACAGTACTTGGGTTTCTGTTGAGA 540
Db 1441 AAGCAGTTAGAGTGTGTTGTCGCAACCTCCGCCAACAGCAGTGGGTTTCTGTTGAGA 1500
Qy 541 GGGTGGACTGAGACAGGAGTACTGGATTTCTAGGCTGACTAAGAAATCCCNAGCCT 600
Db 1501 GGGGGGACTGAGACAGGAAATACTAGATTTCTTAGACCAACTAAGAAATCCCTAAGACT 1560
Qy 601 ANCTGGGAAGTGACCGCATCCATCTTTTAAACATGCGGCTTGCACCTTAGCTCACACCG 660
Db 1561 AGCTGGGAAGTGACCGCTTCCACTTTAAACACCGGGCTTGCACTTAGCTCAGGCCA 1620
Qy 661 ACCAATC-----AGAGAGCTCACTAAATGCTAATCAGGCAAAACAGGAGTAAA 711
Db 1621 ACCAATCAGATACATAAAGAGAGCTCACTAAATGCTAATAGGCAAAACAGGAGATAA 1680
Qy 712 GCAATAGCCATCATCTATTGCTTGAGACGACGCGGAGGACAAGGATGGGATATAA 771
Db 1681 GAAATAGCCATCATCTATTGCTTGAGACGACGCGGAGGACAAGGATGGGATATAA 1740
Qy 772 ACTCAGGCATTCAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 831
Db 1741 ACCAGGCATTCGAGCAGCTACAGCTACCTCTTTGGGTCCCTCCCTTTGTTATGGAG 1800
Qy 832 CTCGTGTTTCACTCTATT 849
Db 1801 CTCGTCTTCACTCTATT 1818

RESULT 28

AAI35980
ID AAI35980 standard; DNA; 1894 BP.

AC AAI35980;

XX 17-OCT-2001 (first entry)

DE Probe #4666 used to measure gene expression in human placenta sample.

XX Probe: microarray; human; placenta; antenatal diagnosis;

XX genetic disorder; ss.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000663.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing

PT gene expression in human placenta.

XX Claim 25; SEQ ID NO 4666; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP).

CC The present sequence is one such probe. The probes are useful for

CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders
XX
SQ Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;

Query Match 53.0%; Score 704.8; DB 4; Length 1894;

Best Local Similarity 89.7%; Pred. No. 3e-220;

Matches 770; Conservative 0; Mismatches 79; Indels 9; Gaps 1;

Qy 1 TCAAAATCGAAGAGCTTTAGACTTTGCTAACCGCCAAAGAGGGGAACTGTTATTTTT 60

Db 961 TCAAAATCGAAGAGCTTTAGACTTTGCTAACCGCCAAAGAGGGGAACTTTTATTTTT 1020

Qy 61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCAATTAATGAGAAGTTAAAGA 120

Db 1021 AGAGGAAAATGCTGTTGTTATGTTAATCAATCGGAATCATCACCGAAGAAGTTAAAGA 1080

Qy 121 AATTTGAGATCGAATATAATGTAGACGAGAGGACTTCAAAACACTGCACCTGGGCGCT 180

Db 1081 AATTCAGAGTCAATATAACGTAGAGCAAAAGAGCTGCNAACACTGGACCTGGGCGCT 1140

Qy 181 CCTCAGCAATGGATGCCCTGGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240

Db 1141 CCTCAGCAATGGATGCCCTGGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1200

Qy 241 TTTACTCTCTTTGGACCTGTATCTTCAACTCTCTTTGTTAAGTTTGTCTTCCAGAAT 300

Db 1201 GTTACTCTCTTTGGACCTGTATCTTAACTCTCTTTGTTAAGTTTGTCTTCCAGAAT 1260

Qy 301 TGAAGCTGTAAGCTACAAATAGTTCCTTCAATGGAAACCCAGATGCATGACTAA 360

Db 1261 CGAAGCAGTAAACTACAAATCGTTTCTCAATGGAGCCCGAGATGCAGTCCATGAGTAA 1320

Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGTCTAGACTATGCTCTGATGTTAATGACATGA 420

Db 1321 AATCTACCGACGACCCCTGGACCGGCTGTCTAGCCATGCTCTGATGTTAATGACATCA 1380

Qy 421 AGTCACCCCTCCGAGGAAATCTCAACTGCAACACCCCTACTACACTCCAATTCAGTAGG 480

Db 1381 AGGCACCCCTCCGAGGAAATCTCAACTGCAACACCTCTACTACGCCCAATTCAGCAGG 1440

Qy 481 AAGCAGTTAGACGAGTTGTGAGCAACCTCCCAACAGTACTTGGGTTTCTGTTGAGA 540

Db 1441 AAGCAGTTAGAGTGGTTGTTGGCCAACTCCCAACAGCAGTGGGTTTCTGTTGAGA 1500

Qy 541 GGGTGGACTGAGACAGGAGTACTGATTTTCTAGCTGACTAAGAAATCCCNAGCCT 600

Db 1501 GGGGGGACTGAGACAGGAAATACTAGATTTCTTAGCCACTAAGAAATCCCTAAGACT 1560

Qy 601 ANCTGGGAAGTGACCGCATCCATCTTTTAAACATGGGGCTTGGAACTTAGCTCACACCG 660

Db 1561 AGCTGGGAAGTGACCGCTTCCACCTTTAAACACCGGGCTTGCAACTTAGCTCAGCCCA 1620

Qy 661 ACCAATC-----AGAGAGCTCACTAAATGCTAATCAGGCAAAACAGGAGTAAA 711

Db 1621 ACCAATCAGATACATAAAGAGAGCTCACTAAATGCTAATTAGGCAAAACAGGAGATAA 1680

Qy 712 GCAATAGCCATCATCTATTGCTTGAGACGACGCGGAGGACAAGGATTTGGGATATAA 771

Db 1681 GAAATAGCCATCATCTATTGCTTGAGACGACGAGGAGGACAATGATCGGGATATAA 1740

Qy 772 ACTCAGGCATTCAGGACGACGACGACGACGACGACGACGACGACGACGACGACGAG 831

Db 1741 ACCAGGCATTCGAGCAGCTACAGCTACCTCTTTGGGTCCCTCCCTTGTATGGAG 1800

Qy 832 CTCGTGTTTCACTCTATT 849

Db 1801 CTCGTCTTCACTCTATT 1818

RESULT 29

ABR45822

ID AC 01-FEB-2002 (first entry)
XX Human breast cell single exon nucleic acid probe #4517.
XX Human; microarray; single exon probe; gene expression; breast; disease;
KW cancer; ss.
XX Homo sapiens.
XX WO200157271-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US0000662.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX New spatially-addressable set of single exon nucleic acid probes, useful
PT for measuring gene expression in sample derived from human breast,
PT comprises number of single exon nucleic acid probes.
XX Claim 1; SEQ ID NO 4517; 327pp + Sequence Listing; English.
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and Br 474 cells. The method involves contacting the
CC probes with a collection of detectably labelled nucleic acids derived
CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC prognosing breast disease. Gene expression analysis is useful for
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a single exon
CC nucleic acid probe of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;
Query Match 53.0%; Score 704.8; DB 4; Length 1894;
Best Local Similarity 89.7%; Pred. No. 3e-220;
Matches 770; Conservative 0; Mismatches 79; Indels 9; Gaps 1;
Qy 1 TCAAAATCGAGAGCTTTAGACTTGGCTTAACCGCCAAAGAGGGGGAACCTGTTATTTT 60
Db 961 TCAAAATCGAGAGCTTTAGACTTGGCTTAACCGCCAAAGAGGGGGAACCTTTTATTTT 1020
Qy 61 AGGGGAAGAATGCTTTAGTATGTTATCAATCTGGAATCATTAATGAGAAAGTTAAAGA 120
Db 1021 AGAGGAAATGCTTTGTTATGTTATCAATCGGAATCATACCGGAGAAAGTTAAAGA 1080
Qy 121 AATTGAGATCGAATATATAATGTTAGAGCAGAGGACCTTCAAAACACCTGCACCCCTGGGCCT 180

Db 1081 AATTCAGGTGGAATATAACGTAGACCAAGAGCTGCAAAACACATGGACCTGGGGCCT 1140
Qy 181 CCTCAGCAATGGATGCCCTGGGACTCTCCCTTTCTTAGGACCTCTAGAGCTATAATATT 240
Db 1141 CCTCAGCAATGGATGCCCTGGGACTCTCCCTTTCTTAGGACCTCTAGAGCTATAATATT 1200
Qy 241 TTTACTCTCTTTGGACCCCTGTATCTCAACTTCCTTCTTAAAGTTTGTCTCTTCCAGAAT 300
Db 1201 GTTACTCTCTTTGGACCCCTGTATCTTAAACCTCTTGTAAAGTTTGTCTCTTCCAGAAT 1260
Qy 301 TGAAGCTGTAAAGCTACAAATAGTCTTCAAAATGGAACCCACAGATGAGTCCATGACTAA 360
Db 1261 CGAAGCAGTAAAACTACAAATGTTCTTCAATGGAGCCCGAGATGAGTCCATGAGTAA 1320
Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGATGTTAATGACATGTA 420
Db 1321 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGATGTTAATGACATGTA 1380
Qy 421 AGTCACCCCTCCGAGGAAATCTCAACTGCAACCCCTACTACACTCCAAATTCAGTAGG 480
Db 1381 AGGCACCCCTCCGAGGAAATCTCAACTGCAACCCCTACTACACTCCAAATTCAGTAGG 1440
Qy 481 AAGCAGTTAGAGCAGTTGTGAGCAACCTCCCAACAGTACTTTGGGTTTCTCTGTTGAGA 540
Db 1441 AAGCAGTTAGAGTGGTTGTTGGCCAACTCCCAACAGTACTTTGGGTTTCTCTGTTGAGA 1500
Qy 541 GGGTGGACTGAGAGACGAGTACTAGTGGATTTCTTAGGCTGACTAAGAAATCCNAAAGCCT 600
Db 1501 GGGGGGACTGAGAGACGAGTAACTAGATTTCTTAGCACTAAGAAATCCCTAAGACT 1560
Qy 601 ANCTGGGAAGGTACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
Db 1561 AGCTGGGAAGGTACCGCTTTCCACCTTTAAACACCGGGCTTGCAACTTAGCTCACGCCCA 1620
Qy 661 ACCAATC-----AGAGAGCTCACTAAATGCTAATCAGGCAAAACACGAGGATAA 711
Db 1621 ACCAATCAGATATAAGAGAGCTCACTAAATGCTAATTAGGCAAAACACGAGGATAA 1680
Qy 712 GCAATAGCCCAATCATCTATTGCTTGAGACAGCAGCGGAAAGGCAAGGATTTGGGATATA 771
Db 1681 GAAATAGCCCAATCATCTATTGCTTGAGACAGCAGGAGGAGCAATGATCGGATATA 1740
Qy 772 ACTCAGCATTAAGCCAGCAACAGCAACCCCTTTGGGTCCTCCCTCCCATTTGATGGAG 831
Db 1741 ACCCAGCATTCGAGCCAGCTACAGTACCTCTTTGGGTCCCTCTCTTTGATGGAG 1800
Qy 832 CTCTGTTTTCACCTATT 849
Db 1801 CTCTGTTTTCACCTATT 1818

RESULT 30
ABA25978
ID ABA25978 standard; DNA; 1894 BP.
XX
XX ABA25978;
XX
XX 23-JAN-2002 (first entry)
DT
DE Probe #4444 for gene expression analysis in human heart cell sample.
XX
XX Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX Homo sapiens.
XX WO200157274-A2.
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US0000666.

XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts.
XX Claim 1; SEQ ID NO 4444; 530pp; English.
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 U; 0 Other;
Query Match 53.0%; Score 704.8; DB 4; Length 1894;
Best Local Similarity 89.7%; Pred. No. 3e-220;
Matches 770; Conservative 0; Mismatches 79; Indels 9; Gaps 1;
QY 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCGAAAGCGGGAAACCTTTTATTTT 60
DB 961 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCGAAAGCGGGAAACCTTTTATTTT 1020
QY 61 AGGGGAAGAATGCTTTAGTATGTTATCAATCTGGAATCATTTACTGAGAAAGTTAAGA 120
DB 1021 AGAGGAAATGCTTTGTTTATGTTATCAATCGGAATCATCACCGAGAAAGTTAAGA 1080
QY 121 AATTGAGATCGAATATAATGTCAGCAGAGGACCTTCAAAACACTGCACCTGGGSCCT 180
DB 1081 AATTCAAGGTCGAATATAATGTCAGCAGAGGACCTTCAAAACACTGCACCTGGGSCCT 1140
QY 181 CCTCAGCCAAATGGATGCCCTGGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATATT 240
DB 1141 CCTCAGCCAAATGGATGCCCTGGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATATT 1200
QY 241 TTATCTCCTCTTTGGACCCCTGTATCTTCAACTTCCTTTGTTAAAGTTCTCTTCCAGAAT 300
DB 1201 GTTACTCCTCTTTGGACCCCTGTATCTTCAACTTCCTTTGTTAAAGTTCTCTTCCAGAAT 1260
QY 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAANTGGACCCCGAGTCAGTCCCATGACTAA 360
DB 1261 CGAAGCGAGTAAAGCTACAAATAGTTCTTCAANTGGACCCCGAGTCAGTCCCATGACTAA 1320
QY 361 AATCTACCGTGGACCCCTGGACCGGCTGTCTAGACTATGCTGTGATGTTAATGACATTGA 420
DB 1321 AATCTACCGACGACCCCTGGACCGGCTGTCTAGCCCATGCTCTGATGTTAATGACATCA 1380
QY 421 AGTCACCCCTCCCGAGGAAATCTCAACTGCACAAACCCCTACTACATCTCCAATTCAGTAGG 480
DB 1381 AGGCACCCCTCCCGAGGAAATCTCAACTGCACAAACCCCTACTACATCTCCAATTCAGTAGG 1440
QY 481 AAGCAGTTAGACGAGTTGTGACGCCAACCTCCGCCAACAGTACTTGGGGTTTCTGTTGAGA 540
DB 1441 AAGCAGTTAGAGTGGTTGTTGGCCAACTCCGCCAACAGTACTTGGGGTTTCTGTTGAGA 1500

QY 541 GGGTGGACTCAGAGACGAGCTAGCTGGATTTCTTAGCTGACTAAGAATCCCNAGCCT 600
DB 1501 GGGGGGACTCAGAGACGAGGAATAACTAGATTTCTTAGACCAACTAAGAATCCCTAAGACT 1560
QY 601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGGAACTTAGCTCACACCG 660
DB 1561 AGCTGGGAAGGTGACCGCTTCCACCTTTAAACACCGGGCTTGCNACTTAGCTCAGGCCA 1620
QY 661 ACCAATC-----AGAGAGCTCACTAAATGCTAATCAGGCAGGAGGATTTGGATATAA 711
DB 1621 ACCAATCAGATACCTAAAGAGAGCTCACTAAATGCTAATAGGCAAAAACAGGAGATAA 1680
QY 712 GCATAGCCAATCATCTATTGCTCAGAGCAGCGGGAAGGACAAGGATTTGGATATAA 771
DB 1681 GAAATAGCCAATCATCTGTTGCTGACAGCAGAGGAGGACAATGATCGGATATAA 1740
QY 772 ACTCAGGCAATTCAGGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGTTATGGAG 831
DB 1741 ACCAGGCAATTCAGGCCAGCTACAGTACCCTCTTTGGGTCCCTCCCTTTGTTATGGAG 1800
QY 832 CTCTGTTTTCACCTCTATT 849
DB 1801 CTCTGTTTTCACCTCTATT 1818

Search completed: January 22, 2005, 16:40:42
Job time : 618.913 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2005, 12:03:14 ; Search time 123.063 Seconds
(without alignments)
7676.074 Million cell updates/sec

Title: US-09-319-156b-12
Perfect score: 1329
Sequence: 1 tcaaatcgaagcgttag.....gcaagatccccagtaaca 1329

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 35539441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1326	99.8	1329	4	US-08-979-847B-108
2	731.8	55.1	2946	3	US-09-175-828-3
3	545	41.0	635	4	US-08-979-847B-102
4	344.6	25.9	443	4	US-09-513-999C-21394
5	302	22.7	1481	4	US-08-979-847B-105
6	265.2	20.0	800	4	US-08-979-847B-127
7	257.8	19.4	764	4	US-08-979-847B-126
8	208	15.7	279	1	US-08-686-878A-50
9	208	15.7	279	1	US-08-721-489-4
10	194.6	14.6	456	4	US-09-621-976-9366
11	166	12.5	1722	3	US-08-691-563C-58
12	166	12.5	1722	4	US-09-374-766-58
13	166	12.5	1722	4	US-08-979-847B-54
14	165	12.4	1859	3	US-08-691-563C-46
15	165	12.4	1859	4	US-09-374-766-46
16	161.8	12.2	80246	3	US-08-979-847B-42
17	161.8	12.2	80246	4	US-09-078-294-4
18	151.8	11.4	80595	3	US-09-078-294-3
19	123.6	9.3	506	4	US-09-621-976-2599
20	107.8	8.1	3910	3	US-09-120-653D-1
21	103.2	7.8	149	4	US-09-513-999C-24228
22	85.8	6.5	2908	4	US-09-799-451-249
23	80.6	6.1	1704	1	US-08-007-282B-1
24	80.6	6.1	2518	3	US-09-011-745-1
25	80.6	6.1	4776	4	US-09-555-352-1
26	80.6	6.1	5865	3	US-09-011-745-8
27	78.2	5.9	464	3	US-09-120-653D-6

28	76	5.7	492	4	US-09-621-976-11057	Sequence 11057, A
29	73.4	5.5	442	4	US-09-621-976-17982	Sequence 17982, A
30	69.6	5.2	2809	4	US-09-620-312D-171	Sequence 171, Appl
31	68	5.1	2342	4	US-09-904-615-12	Sequence 12, Appl
32	57	4.3	7218	1	US-08-232-463-14	Sequence 14, Appl
33	50.8	3.8	8060	3	US-08-766-528-1	Sequence 1, Appli
34	50.8	3.8	8060	4	US-09-661-858-1	Sequence 1, Appli
35	49.2	3.7	8132	3	US-08-766-528-3	Sequence 3, Appli
36	49.2	3.7	8132	4	US-09-661-858-3	Sequence 3, Appli
37	48.4	3.6	304	4	US-09-513-999C-26710	Sequence 26710, A
38	47.8	3.6	502	4	US-09-621-976-9610	Sequence 9610, Ap
39	46.4	3.5	1911	1	US-08-258-420-8	Sequence 8, Appli
40	46.4	3.5	1965	1	US-08-258-420-9	Sequence 9, Appli
41	46.4	3.5	3925	3	US-09-011-745-9	Sequence 9, Appli
42	46.4	3.5	8202	1	US-08-258-420-13	Sequence 13, Appl
43	45.2	3.4	342	4	US-09-513-999C-13878	Sequence 13878, A
44	44.8	3.4	6028	3	US-09-011-745-5	Sequence 5, Appli
45	44.6	3.4	1989	3	US-09-232-278A-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1

US-08-979-847B-108

; Sequence 108, Application US/08979847B

; Patent No. 6582703

; GENERAL INFORMATION:

APPLICANT: PERON, HERVE
BESEME, FREDERIC
BEDIN, FREDERIC
PARAMHOS-BACCALA, GLAUCIA
KOMURIAN-PRADEL, FLORENCE
JOLIVET-REYNAUD, COLETTE
MANDRAND, BERNARD
GARSON, JEREMY
TURE, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210

CORRESPONDENCE ADDRESS:

ADDRESSEE: OLIFF & BERRIDGE, PLC

STREET: P.O. BOX 19928

CITY: ALEXANDRIA

STATE: VA

COUNTRY: USA

ZIP: 22320

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/979,847B

FILING DATE: 26-No. 6582703-1997

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BERRIDGE, WILLIAM P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: WPB 39046A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 108:

SEQUENCE CHARACTERISTICS:

LENGTH: 1329 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 108:

US-08-979-847B-108

Query Match 99.8%; Score 1326; DB 4; Length 1329;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAAATCGAAGAGCTTTAGACTGCTTAACCGCCAAAGAGGGGGAACCTGTTTATTTT 60
DB 1 TCAAAATCGAAGAGCTTTAGACTGCTTAACCGCCAAAGAGGGGGAACCTGTTTATTTT 60

QY 61 AGGGGAAGATGCTGTTAGTATGTTAATCAATCTGGAATCAATTAATCTGAGAAAGTTAAAGA 120
DB 61 AGGGGAAGATGCTGTTAGTATGTTAATCAATCTGGAATCAATTAATCTGAGAAAGTTAAAGA 120

QY 121 AATTTGAGATCGAATAATAATGTAGAGCAGAGGACCTTCAAAAACATGCAACCTGGGGCCT 180
DB 121 AATTTGAGATCGAATAATAATGTAGAGCAGAGGACCTTCAAAAACATGCAACCTGGGGCCT 180

QY 181 CCTCAGCAATGGATGCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATTT 240
DB 181 CCTCAGCAATGGATGCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATTT 240

QY 241 TTTACTCCTCTTTGGACCCCTGATCTTCAACTTCTTGTAAAGTTTGTCTTTCCAGAA 300
DB 241 TTTACTCCTCTTTGGACCCCTGATCTTCAACTTCTTGTAAAGTTTGTCTTTCCAGAA 300

QY 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAACCCAGATGCAGTCCATGACTAA 360
DB 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAACCCAGATGCAGTCCATGACTAA 360

QY 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGATGTTAATGACATGA 420
DB 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGATGTTAATGACATGA 420

QY 421 AGTCACCCCTCCGAGGAATCTCAACTGACACACCCCTACTACACTCCAAATTCAGTAGG 480
DB 421 AGTCACCCCTCCGAGGAATCTCAACTGACACACCCCTACTACACTCCAAATTCAGTAGG 480

QY 481 AAGCAGTTAGAGCAGTTGTGAGCCAACTCTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 540
DB 481 AAGCAGTTAGAGCAGTTGTGAGCCAACTCTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 540

QY 541 GGGTGGACTGAGACAGGACTAGCTGGATTTCTAGGCTGACTAAGAATCCNAGCCT 600
DB 541 GGGTGGACTGAGACAGGACTAGCTGGATTTCTAGGCTGACTAAGAATCCNAGCCT 600

QY 601 ANCTGGGAAGTGACCGCATCTTTTAAACATGGGGCTTGCAACTTAGCTCACACCG 660
DB 601 ANCTGGGAAGTGACCGCATCTTTTAAACATGGGGCTTGCAACTTAGCTCACACCG 660

QY 661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAAACAGGAGTTAAAGCAATAGCC 720
DB 661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAAACAGGAGTTAAAGCAATAGCC 720

QY 721 AATCATCTATTGCTGAGACAGACGGGGAAGGACAGGATTTGGGATATAAATCAGGCA 780
DB 721 AATCATCTATTGCTGAGACAGACGGGGAAGGACAGGATTTGGGATATAAATCAGGCA 780

QY 781 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCAATGTTATGGGAGCTCTGTTT 840
DB 781 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCAATGTTATGGGAGCTCTGTTT 840

QY 841 CACTCTATTTCACTCTAATTAATCATGCAACTGCACTCTTCTGCTCGGTGTTTATGG 900
DB 841 CACTCTATTTCACTCTAATTAATCATGCAACTGCACTCTTCTGCTCGGTGTTTATGG 900

QY 901 CTCAAGCTGAGCTTTGTTGGCCATCCACACTGCTGTTTGGCCACCGTCAAGACCCGCT 960
DB 901 CTCAAGCTGAGCTTTGTTGGCCATCCACACTGCTGTTTGGCCACCGTCAAGACCCGCT 960

QY 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCACACTGTGCTCCTGATCCAGAGGT 1020
DB 961 GCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCACACTGTGCTCCTGATCCAGAGGT 1020

RESULT 2

US-09-175-928-3
; Sequence 3, Application US/09175928A
; Patent No. 6312921
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Werberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Mi, Sha
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 6006B.AJ172A
; CURRENT APPLICATION NUMBER: US/09/175,928A
; CURRENT FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2946
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-175-928-3

Query Match 55.1%; Score 731.8; DB 3; Length 2946;
Best Local Similarity 90.7%; Pred. No. 7.7e-240;
Matches 794; Conservative 0; Mismatches 69; Indels 12; Gaps 1;

QY 1 TCAAAATCGAAGAGCTTTAGACTGCTTAACCGCCAAAGAGGGGGAACCTGTTTATTTT 60
DB 2067 TCAAAATCGAAGAGCTTTAGACTGCTTAACCGCTGAAAGAGGGGGAACCTGTTTATTTT 2126

QY 61 AGGGGAAGATGCTGTTAGTATGTTAATCAATCTGGAATCAATTAATCTGAGAAAGTTAAAGA 120
DB 2127 AGGGGAAGATGCTGTTATTTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2186

QY 121 AATTTGAGATCGAATAATAATGTAGAGCAGAGGACCTTCAAAAACATGCAACCTGGGGCCT 180
DB 2187 AATTCGAGATCGAATAACAGTAGACAGAGGAGCTTCGAAACACTGACCTGGGGCCT 2246

QY 181 CCTCAGCAATGGATGCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATTT 240
DB 2247 CCTCAGCAATGGATGCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATTT 2306

QY 241 TTTACTCCTTTTGGACCCCTGATCTTCAACTTCTTGTAAAGTTTGTCTCTTCCAGAA 300
DB 241 TTTACTCCTTTTGGACCCCTGATCTTCAACTTCTTGTAAAGTTTGTCTCTTCCAGAA 300

Db 2307 GCTACTCCTCTTTGGACCCCTGTATCTTTAACTCCTTGTAACTTTGTCTCTTCCAGAT 2366
Qy 301 TGAAGCTGTAAAGCTACAAATAGTCTTCTTAAATGGAAACCCAGATGCGTCCAGCTAA 360
Db 2367 CGAAGCTGTAAACTA-----CAATGGAGCCCAAGATGCGTCCAGCTAA 2414
Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGATGTTAATGACATTGA 420
Db 2415 GATCTACCGCAGACCCCTGGACCGGCTGTAGCCCAAGATCTGATGTTAATGACATCA 2474
Qy 421 AGTCACCCCTCCCGAGGAATCTCAATGTCACACCCCTACTACTACTCAATTCAGTAGG 480
Db 2475 AGGCACCCCTCTCGAGGAATCTCAGCTGCAACCTCTACTACTAGCCCAATTCAGAGG 2534
Qy 481 AAGCAGTAGAGAGTGTGAGCAACCTCCCAACAGTACTTTGGGTTTCTGTTTGA 540
Db 2535 AAGCAGTAGAGCGGTGCTGGGCCAACCTCCCAACAGCACTTAGGTTTCTGTTTGA 2594
Qy 541 GGTGAGTGTAGACAGACAGGACTAGTGGATTTCTTAGGCTGACTAAGATCCCAAGCCT 600
Db 2595 TGGGGGACTGAGACAGGACTAGTGGATTTCTTAGGCTGACTAAGATCCCAAGCCT 2654
Qy 601 ANCTGGGAAGTGCACGCTATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
Db 2655 AGCTGGGAAGTGCACATCTCACCTTTAAACAGGGGCTTGCAACTTAGCTCACACCTG 2714
Qy 661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAACAGGAGGTAAAGCAATAGCC 720
Db 2715 ACCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAAACAGGAGGTAAAGCAATAGCC 2774
Qy 721 AATCATCTATTGCTGAGACAGCGGGAAGCAAGGATTTGGGATATAAATCTCAGGCA 780
Db 2775 AATCATCTATTGCTGAGACAGCAGGAGGAGCAATGATCGGATATAAACCAGTC 2834
Qy 781 TTCAAGCCGACACAGCAACCCCTTTGGTCCCTCCCTATTCATGAGGAGCTCTGTTT 840
Db 2835 TTGAGCGGCAACGGAACCCCTTTGGTCCCTCCCTTTGATGGAGGCTCTGTTT 2894
Qy 841 CACTCTATTCTACTATTAAATCATGCAACTGCA 875
Db 2895 CATGCTATTCTACTATTAAATCTTGCAACTGCA 2929

RESULT 3

US-08-979-847B-102
; Sequence 102, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:

APPLICANT: PERRON, HERVE
BESEME, FREDERIC
BEDIN, FREDERIC
PARANHOS-BACCALA, GLAUCIA
KOMURIAN-PRADEL, FLORENCE
JOLIVET-REYNAUD, COLETTE
MANDRAND, BERNARD
GARSON, JEREMY
TUKU, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P. O. BOX 19928
CITY: ALEXANDRIA
STATE: VA

COUNTRY: USA
ZIP: 22320

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-NO. 6582703-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 635 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-08-979-847B-102

Query Match 41.0%; Score 545; DB 4; Length 635;
Best Local Similarity 92.4%; Pred. No. 3.1e-176;
Matches 572; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 257 CCTGTATCTTCAACTGTCCTTGTAAAGTTTGTCTTCCAGATTTGAAGCTGTAAGCTA 316
Db 1 CCTGTATCTTAACTCCTCTTGAAGTTTGTCTTCCAGATCAAACTGTAAACTA 60
Qy 317 CAAATGTTCTTCAATGGAACCCAGATGTCAGTCCATGACTAAATCTACCGTGGACCC 376
Db 61 CAAATGTTCTTCAATGGAACCCAGATGTCAGTCCATGACTAAATCTACCGTGGACCC 120
Qy 377 CTGGACCGGCTGCTAGACTATGCTCTGATGTTAATGACATTTGAAGTCAACCTCCCGAG 436
Db 121 CTGGACCGGCTGCTAGACTGCTGATGTTAATGACATTTGAAGTCAACCTCCCGAG 180
Qy 437 GAAATCTCAACTGTCACAAACCCCTTACTACTCAATTCAGTAGGAGGAGTTCAGAGCT 496
Db 181 GAAATCTCAACTGTCACAAACCCCTTACTACTCAATTCAGTAGGAGGAGTTCAGAGCT 240
Qy 497 TGTGACGCAACCTCCCAACAGTACTTGGGTTTCTGTTGAGGAGGAGTTCAGAGCT 556
Db 241 CATGACCAACCTCCCAACAGTACTTGGGTTTCTGTTGAGGAGGAGTTCAGAGCT 300
Qy 557 AGGACTAGCTGGATTTCTTAGGCTGACTAAGATCCCAAGCTTANCTGGGAAGTGAAC 616
Db 301 AGGACTAGCTGGATTTCTTAGGCTGACTAAGATCCCAAGCTTANCTGGGAAGTGAAC 360
Qy 617 GCATCCATCTTTAAACATGGGCTTGCAACTTAGCTCACACCGCAATCAGAGAGCTC 676
Db 361 GCATCCATCTTTAAACATGGGCTTGCAACTTAGCTCACACCGCAATCAGAGAGCTC 420
Qy 677 ACTAAATGCTATCAGGCAAAACAGGAGGTAAAGCAATAGCCATCATCTATTGCTG 736
Db 421 ACTAAATGCTATCAGGCAAAACAGGAGGTAAAGCAATAGCCATCATCTATTGCTG 480
Qy 737 AGAGCACAGCGGAGGACAGGATTTGGGATATAAACTCAGGCAATCAAGCCAGCAACAG 796
Db 481 AGAGCACAGCGGAGGACAGGATTTGGGATATAAACTCAGGCAATCAAGCCAGCAACAG 540
Qy 797 CAAACCCCTTTGGGTCCTCCCTCCCATTTGATGGAGCTCTGTTTCACTCTATTTCATCT 856
Db 541 CAAACCCCTTTGGGTCCTCCCTCCCATTTGATGGGCGCTCTGTTTCACTCTATTTCATCT 600
Qy 857 ATTAAATCATGCAACTGCA 875
Db 601 ATTAAATCTTGCAACTGAA 619

RESULT 4

US-09-513-999C-21394
; Sequence 21394, Application US/09513999C

APPLICANT: PERRON, HERVE
BESEME, FREDERIC
BEDIN, FREDERIC
PARANHOS-BACCALA, GLAUCIA
KOMURIAN-PRADEL, FLORENCE
JOLIVET-REYNAUD, COLETTE
MANDRAND, BERNARD
GARSON, JEREMY
TUKU, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-No. 6582703-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 127:
SEQUENCE CHARACTERISTICS:
LENGTH: 800 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 127:
US-08-979-847B-127

Query Match 20.0%; Score 265.2; DB 4; Length 800;
Best Local Similarity 88.2%; Pred. No. 3.3e-80;
Matches 299; Conservative 0; Mismatches 39; Indels 1; Gaps 1;
QY 992 TGTCCACTGTGCTCCTGATCCAGC-GAGGTACCATTGCCACTCCCGATCAGGCTAAAGG 1050
DB 1 TGTCCGCTGTGCTCCTGATCCAGCAGCGGCCATTGCTCTCCCAATTGGGCTAAAGG 60
QY 1051 CTTGCCATTGTTCTGTCATGGCTTAAGTCCCTGGTTGCTTCTTAATAGAACTGAACACTGG 1110
DB 61 CTTGCCATTGTTCTGTCATGGCTTAAGTCCCTGGTTGCTTCTTAATAGAACTGAACACTAG 120
QY 1111 TCACTGGGTTCCAGGTTCTTCTTCCATGACCCATGCGCTTCTTAATAGAACTGAACACTCA 1170
DB 121 TCACTGGGTTCCAGGTTCTTCTTCCATGACCCATGCGCTTCTTAATAGAACTGAACACTCA 180
QY 1171 CCGCATGGCCCAAGATTCCATTCTTTGGTATCTGTGAGGCCAAGAACCCCGAGTCAAGAG 1230
DB 181 CTGCATGGTCCAAAGATTCCATTCTTTGGATCTGTGAGAACCAAGAACCCCGAGTCAAGAG 240
QY 1231 AAGTGGCTTGGCAACCATTTGGGAAGTGGCCCACTGCCATTTTGGTGGGCGCCAC 1290
DB 241 ACACAAGGCTTGGCAACCATTTTGGGAAGTGGCCCACTGCCATTTTGGTGGGCGCCAC 300
QY 1291 CATCTTGGGAGCTGTGGAGCAAGGATCCCGCAAGTAA 1329
DB 301 TATCTTGGGAGCTGTGGAGCAAGGATCCCGCAAGTAA 339

RESULT 7

US-08-979-847B-126
Sequence 126, Application US/08979847B
Patent No. 6582703
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
BESEME, FREDERIC
BEDIN, FREDERIC
PARANHOS-BACCALA, GLAUCIA
KOMURIAN-PRADEL, FLORENCE
JOLIVET-REYNAUD, COLETTE
MANDRAND, BERNARD
GARSON, JEREMY
TUKU, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-No. 6582703-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 126:
SEQUENCE CHARACTERISTICS:
LENGTH: 764 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 126:
US-08-979-847B-126

Query Match 19.4%; Score 257.8; DB 4; Length 764;
Best Local Similarity 88.2%; Pred. No. 1.1e-77;
Matches 291; Conservative 0; Mismatches 38; Indels 1; Gaps 1;
QY 992 TGTCCACTGTGCTCCTGATCCAGC-GAGGTACCATTGCCACTCCCGATCAGGCTAAAGG 1050
DB 1 TGTCCGCTGTGCTCCTGATCCAGCAGCGGCCATTGCTCTCCCAATTGGGCTAAAGG 60
QY 1051 CTTGCCATTGTTCTGTCATGGCTTAAGTCCCTGGTTGCTTCTTAATAGAACTGAACACTGG 1110
DB 61 CTTGCCATTGTTCTGTCATGGCTTAAGTCCCTGGTTGCTTCTTAATAGAACTGAACACTAG 120
QY 1111 TCACTGGGTTCCAGGTTCTTCTTCCATGACCCAGGCTTCTTAATAGAACTGAACACTCA 1170
DB 121 TCACTGGGTTCCAGGTTCTTCTTCCATGACCCAGGCTTCTTAATAGAACTGAACACTCA 180
QY 1171 CCGCATGGCCCAAGATTCCATTCTTTGGTATCTGTGAGGCCAAGAACCCCGAGTCAAGAG 1230
DB 181 CTGCATGGTCCAAAGATTCCATTCTTTGGATCTGTGAGAACCAAGAACCCCGAGTCAAGAG 240

QY 1231 ANGTGAGGCTTGCACCACTTTGGAAAGTGGCCCACTGCCATTTTGTAGCGGCCCAAC 1290
Db 241 ACACAGGCTTGCACCACTTTGGAAAGTGGCCCACTGCCATTTTGTAGCGGCCCAAC 300
QY 1291 CATCTTGGGAGCTGTGGGACCAAGGATCCC 1320
Db 301 TATCTTGGGAGCTGTGGGACCAAGGATCCC 330

RESULT 8
US-08-686-878A-50
; Sequence 50, Application US/08686878A
; Patent No. 5708157
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Evans, Cheryl
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/686,878A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-686-878A-50

Query Match 15.7%; Score 208; DB 1; Length 279;
Best Local Similarity 85.1%; Pred. No. 6.5e-61;
Matches 223; Conservative 1; Mismatches 38; Indels 0; Gaps 0;
QY 614 ACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAG 673
Db 1 RCCACATCCACCTTTAAACACGGGGNTTGCAANAAGATNACACTTTGACCAATCAGAGAG 60
QY 674 CTCACATAAATGCTAATCAGGCAAAAACAGGAGGTAAGCAATAGCAATCATCTATTGC 733
Db 61 NTCANTAAATGATNATTNGGCAAAAACAGGAGGTAAGCAATAGCAATCATCTATTGC 120
QY 734 CTGAGAGCACAGCGGGAGGCAAGGATTTGGGATATAAACTCAGGCAATTCAGGCCAACAA 793
Db 121 CTGAGAGCACAGGAGGACATATCGGGATATAAACCAGATTTTNGAGCGCGCAA 180
QY 794 CAGCAACCCCTTTGGGTCCCTCCCATTTGTATGGGAGCTCTGTTTTCATCTATTTCAC 853

Db 181 CGSCAACCCCTTTGGGTCCCTCCCATTTGTATGGGAGCTNTGTTTTCATCTATTTCAN 240
QY 854 TCTATTAAATCATGCAACTGCA 875
Db 241 TATATTAAATNTTGCACACTGCA 262
RESULT 9
US-08-721-489-4
; Sequence 4, Application US/08721489
; Patent No. 5786465
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/721,489
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-721-489-4

Query Match 15.7%; Score 208; DB 1; Length 279;
Best Local Similarity 85.1%; Pred. No. 6.5e-61;
Matches 223; Conservative 1; Mismatches 38; Indels 0; Gaps 0;
QY 614 ACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAG 673
Db 1 RCCACATCCACCTTTAAACACGGGGNTTGCAANAAGATNACACTTTGACCAATCAGAGAG 60
QY 674 CTCACATAAATGCTAATCAGGCAAAAACAGGAGGTAAGCAATAGCAATCATCTATTGC 733
Db 61 NTCANTAAATGATNATTNGGCAAAAACAGGAGGTAAGCAATAGCAATCATCTATTGC 120
QY 734 CTGAGAGCACAGCGGGAGGCAAGGATTTGGGATATAAACTCAGGCAATTCAGGCCAACAA 793
Db 121 CTGAGAGCACAGGAGGACATATCGGGATATAAACCAGATTTTNGAGCGCGCAA 180
QY 794 CAGCAACCCCTTTGGGTCCCTCCCATTTGTATGGGAGCTCTGTTTTCATCTATTTCAC 853
Db 181 CGSCAACCCCTTTGGGTCCCTCCCATTTGTATGGGAGCTNTGTTTTCATCTATTTCAN 240

Qy 854 TCATTAAATCATGCAACTGCA 875
Db 241 TTTATTAAATTTGCAACTGCA 262

RESULT 10

US-09-621-976-9366/c
; Sequence 9366, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054P22
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 9366
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-9366

Query Match 14.6%; Score 194.6; DB 4; Length 456;
Best Local Similarity 84.2%; Pred. No. 3.6e-56;
Matches 262; Conservative 0; Mismatches 26; Indels 23; Gaps 3;

Qy 548 CTGAGAGACAGGACTAGCTGGATTTCTAGGCTGACTAAGAAATCCNAAGCCTTACTGGG 607
Db 307 CTGAGAGACAGGACTAGCTGGATTTCTAGGCGGACTAAGAAATCCCTAAGCCTAGCTGGG 248
Qy 608 AAGGTGACCCATCATCTTTAAACATGGGCTTTGCAACTTAGTCAACCCGACCAATC 667
Db 247 AAGGTGACCAATCCACTTTAGCAGGGGCTTTGCAACTTAGTCAACCCGACCAATC 188
Qy 668 -----AGAGAGCTCACTAAATGCTAATCAGGCAAAACAGGAGTAAAGCAATAG 718
Db 187 AGGTAGTAAAGCGAGCTCACTAAATGCTAATAGGCTAAACAGGAGGTAAAC-CAATAG 129
Qy 719 CCAATCATCTATTGCTGAGAGCAGCGGAGGAGCAAGGATGGGATATAAATCAGG 778
Db 128 CCAATCATCTATTCTCTCAGAGCAGCAGGAGGAGCAATGATAGGAGTATAAAC----- 74
Qy 779 CATTCAAGCCAGCAACAGCAACCCCTTTGGGCTCCCTCCATTTGATGGAGCTCTGTT 838
Db 73 -----CTGGCAACGGCTGCTCTGGGCTCCCTCTCTGGGCTCCCTCTGTTGATGGAGCTCTGTT 22
Qy 839 TTCACCTATT 849
Db 21 TTCACCTATT 11

RESULT 11

US-08-691-563C-58
; Sequence 58, Application US/08691563C
; Patent No. 6001987
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glauca PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge

; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/691.563C
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38588
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1722 base pairs
; TYPE: nucleotide
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-691-563C-58

Query Match 12.5%; Score 166; DB 3; Length 1722;
Best Local Similarity 60.5%; Pred. No. 5.9e-46;
Matches 333; Conservative 0; Mismatches 200; Indels 17; Gaps 3;

Qy 2 CAAAATCGAAGAGCTTTAGACTTTGCTAAACCGCAAAAGAGGGGGAACCTGTTTATTTTAA 61
Db 547 CAAAACGGCGAGGCCACACCTCTCTGCTGAGAAAGAGGAGCTCTGACCTTCTTA 606
Qy 62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTAATGAGAAAGTTAAAGAA 121
Db 607 GGGGAAGAATGTTGTTTATACACTAACCACTAGTACGGATAGTACGAGATGCCCTGGCAT 666
Qy 122 ATTTGAGATCGAATATAATGATAGCAGAGGACCTTCAAAACACTGCACCTGGGGCTC 181
Db 667 TTACAGAAAGGGCTTCTGATATCAGACAAATGCTTTCAAACTCTTATACAA-----C 720
Qy 182 CTGAGCAATGATGCGCTGGACTCTCCCTCTCTTAGGACCTCTAGCAGCTATAATATT 241
Db 721 CTCTGGAGTTGGCAACATGGCTTCTTCCATTTCTAGTCCCATGGCAGCATCTTGCTG 780
Qy 242 TTAATCTCTTTGGACCTGATCTTCAACTCTCTTGTAAAGTTTGTCTCTTCAGAAAT 301
Db 781 TTAATCACTTTGGGCTGATTTTAAAGCTTTCTTGTCAAAATTTGTTCTCTTAGGATC 840
Qy 302 GAAGCTGTAAGCTACAAATAGTTCTCAAAATGGAACCCAGATCGAGTCCATGACTAAA 361
Db 841 GAAGCAATCAAGCTACAGATGGTCTTCAAAATGGAACCCCAATAGATTTCACTAACAC 900
Qy 362 ATCTAGCTGGACCCCTGGACCGCTGCTAGACTATGCTCTGATGTTAATGACTTAA 421
Db 901 TTCTACAGGACCCCTGGACGATCCACTGGGACCTTCCACT-----AGCTAGA 950
Qy 422 GTCAACCCCTCCGAGAAATCTCAACTGCAACACCCCTACTACACTCCAAATTCAGTAGGA 481
Db 951 GATTCCCTCTCGAAGACACTACAACTGCAAGGCGCCCTCTTTGCGCCCTATCCAGCAGGA 1010
Qy 482 AGCAGTAGCAGGATTTGTGAGCCAACTCCCAACAGTACTCTGGTTTCTCTCTGAGAG 541
Db 1011 AGTAGCTAGACCGGTTCATCGGCCAA-ATTCCCAACAGCAGTGTGGGGTGTCTCTGTTAGAG 1069
Qy 542 GGTGGACTGA 551
Db 1070 GGGGGATTGA 1079

RESULT 12
US-09-374-766-58
; Sequence 58, Application US/09374766
; Patent No. 6579526
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glauca PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/374,766
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,563
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38588
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1722 base pairs
; TYPE: nucleotide
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-374-766-58

Query Match 12.5%; Score 166; DB 4; Length 1722;
Best Local Similarity 60.5%; Pred.No. 5.9e-46;
Matches 333; Conservative 0; Mismatches 200; Indels 17; Gaps 3;

QY 2 CAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAGAGGGGAACTGTTTATTTTAA 61
DB 547 CAAACCCCGGCGCCACACTCTCTACTGCTGAGAAAGGAGACTCTGCACCTTCTTA 606
QY 62 GGGGAAGAATGCTGTAGTATGTTAATCAATCTGGAATCATCTAGAGAAAGTTAAGAA 121
DB 607 GGGGAAGAAGTGTGTTTTTACACTAACCAAGTCAGGGATAGTACGAGATGCCACCTGGCAT 666
QY 122 ATTTGAGATCGAATATATGATAGACAGAGACCTTCAAAACACTGCAACCTGGGCGCTC 181
DB 667 TTACAGGAAGGGCTTCTGATATACAGCAATGCCCTTCAAACCTCTATACAA-----C 720
QY 182 CTCAGCCAAATGATGCCCTGGACTCTCCCTTTCTTAGGACCTCTAGAGCTATATATTT 241
DB 721 CTCTGGAGTTGGCAACATGGCTTCTTCCATTTCTAGTCCCATGGCAGCCATCTTGCTG 780
QY 242 TTACTCTCTTTGGACCCCTGATCTTCAACTTCCTTGTAACTTGTCTCTTCCGAAATT 301

DB 781 TTACTCACCTTTGGGCCCTGTATTTTAAAGCTTCTTGTCAAATTTGTTTCTCTAGGATC 840
QY 302 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAAACCCAGATGCGAGTCCATGACTAAA 361
DB 841 GAAGCCATCAAGCTACAGATGCTTTACAAATGGAAACCCCAAATGAGTTCAAACCTAACAC 900
QY 362 ATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGTTAATGACATTGAA 421
DB 901 TTCTACCAAGGACCCCTGGAAACGATCCACTGGCACTTCCACT-----AGCCTAGA 950
QY 422 GTACACCCCTCCGAGGAAATCTCAACTGCACAACCCCTACTACACTTCCAAATTCAGTAGGA 481
DB 951 GATTCCTCTCTGGAAGACACTACAACTGACGAGGCGCCCTCTTTGCCCTTATCCAGCAGGA 1010
QY 482 AGCAGTTAGACAGATTGTGACGCAACCTCCCAACAGTACTTGGGTTTCTCTGTTGAGAG 541
DB 1011 AGTAGCTAGAGCGGTCATCGGCAA-ATTCCCAACAGCAGTTGGGGTGTCTCTGTTTAGAG 1069
QY 542 GTGGAGTGA 551
DB 1070 GGGGGATTGA 1079

RESULT 13
US-08-979-847B-54
; Sequence 54, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; BESEME, FREDERIC
; BEDIN, FREDERIC
; PARANHOS-BACCALA, GLAUCIA
; KOMURIAN-PRADEL, FLORENCE
; JOLIVET-REYNAUD, COLETTE
; MANDRAND, BERNARD
; GARSON, JEREMY
; TURE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,847B
; FILING DATE: 26-No. 6582703-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BERRIDGE, WILLIAM P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 39046A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1722 base pairs
; TYPE: nucleotide
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; SEQUENCE DESCRIPTION: SEQ ID NO: 54:
US-08-979-847B-54

Query Match 12.58; Score 166; DB 4; Length 1722;
Best Local Similarity 60.5%; Pred. No. 5.9e-46;
Matches 333; Conservative 0; Mismatches 200; Indels 17; Gaps 3;

QY 2 CAAAATCGAAGCTTTAGACTTGTCTAAGCCGCAAGAGAGGGGAACCTGTTTATTTTA 61
DB 547 CAAAACGCGAGCCCAACCTCTCTGCTGAGAAAGAGAGACTCTGCACCTCTTAA 606

QY 62 GGGGAAGAATGCTGTGTAGTATGTTAAATCAATCTGGAATATTACTGAGAAAGTTAAAGAA 121
DB 607 GGGGAAGAGTGTGTTTACACTTAACCACTGAGGATAGTACGAGATGCCACTGGCAT 666

QY 122 ATTGAGATCGAATATAATCTAGAGCAGAGACCTTCAAAACACTGCACCTGGGGCTC 181
DB 667 TTACAGGAAGAGGCTTCTGATATCAGAAATGCTTTCAAACTCTTATACCAA-----C 720

QY 182 CTGAGCCAAATGGATGCTCGACTCTCCCTCTCTTAGGACCTCTAGCAGCTATAATATT 241
DB 721 CTCTGGAGTTGGCAACATGGCTTCTTCAATTTCTAGTCCCATGGCAGCATCTTGCTG 780

QY 242 TTAATCTCTTTGGACCTGTATCTTCAACTTCTCTGTTAAAGTTTGTCTTCCAGAAAT 301
DB 781 TTAATCACTTTGGGCTCTGTTTAAAGCTTCTTGTCAAAATTTGTTCTCTAGGATC 840

QY 302 GAAGCTGTAAGCTACAAATAGTCTTCAATGGAACCCAGATGCGATGCATGACTAAA 361
DB 841 GAAGCCATCAAGCTACAGATGCTTTACAAATGGAACCCCAAAATGAGTTTCAACTAAC 900

QY 362 ATCTACGCTGACCCCTGGACCGCTGCTAGACTGCTCTGATGTTAATGACATTGAA 421
DB 901 TTCTACCAAGAGCCCTGGACAGATCCACTGGACATTTCCACT-----AGCTAGA 950

QY 422 GTCAACCCCTCCGAGGAATCTCAACTGCACAAACCCCTACTACACTCCAAATTCAGTAGGA 481
DB 951 GATTCCCTCTGGAAGACACTACAACTGCGAGGCGCTTCTTGGCCCTATCCAGCAGGA 1010

QY 482 AGCAGTTAGAGCTGTTGCTAGCCAACTCCGCCAAGACTGTTGGGTTTCTGTTGAGAG 541
DB 1011 AGTAGCTAGAGCGGTCTATCGGCCAA-ATTCCCAACAGCAGTTGGGGTGTCTGTTAGAG 1069

QY 542 GGTGGACTGA 551
DB 1070 GGGGGATTGA 1079

RESULT 14
US-08-691-563C-46
; Sequence 46, Application US/08691563C
; Patent No. 6001987
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glauca PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C
FILING DATE: 02-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 38588
TELEPHONE: 703-836-2787
TELEFAX: 703-836-6400
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1859 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-691-563C-46

Query Match 12.44; Score 165; DB 3; Length 1859;
Best Local Similarity 58.3%; Pred. No. 1.4e-45;
Matches 369; Conservative 0; Mismatches 247; Indels 17; Gaps 4;

QY 2 CAAAATCGAAGCTTTAGACTTGTCTAAGCCGCAAGAGAGGGGAACCTGTTTATTTTA 61
DB 1197 CAAAACGCTCAAGCCCTAGACCTCTCTGCTGAGAAAGAGGAGCTCTGCACCTCTTA 1256

QY 62 GGGGAAGAATGCTGTGTAGTATGTTAAATCAATCTGGAATCAATCTGAGAAAGTTAAAGAA 121
DB 1257 AGGGAAGATGTTGTTCTTACACTAACCACTAGGGATAGTA--TGAGATGCTGCCGGC 1314

QY 122 ATTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCTGGGGCTC 181
DB 1315 ATTACAGAAAAGAGGCTTCTGAAATCAGACACGCTTTTCAAAAT-----CTATACCAAC 1370

QY 182 CTCAGCCAAATGGATGCTCGACTCTCCCTCTCTTAGGACCTCTAGCAGCTATAATATT 241
DB 1371 CTCTGAGTTGGGCAACATGGTTTCTTCTATGTCCTCCATGGCTGCCATCTTGCTA 1430

QY 242 TTAATCTCTTTGGACCTGTATCTTCAACTCTCTTGTAAAGTTTGTCTCTCAGAAAT 301
DB 1431 TTAATCGCTTTGGGCTGTATTTTAACTCTTGTCAAATTTGTTTCTTAGATC 1490

QY 302 GAAGCTGTAAGCTACAAATAGTCTTCAAAATGGAACCCAGATGCGATGCATGACTAAA 361
DB 1491 GAGGCCATCAAGCTACAGATGCTTTACAAATGGAACCCCAAAATGAGCTCAACTATCAAC 1550

QY 362 ATCTACGCTGGACCCCTGGACCGCTGCTAGACTATGCTCTGATGTTAATGACATTGAA 421
DB 1551 TTCTACTGAGGACCCCTAGACCAACCCCTTTCCTGCTTCAAAAGAGTT-----1605

QY 422 GTCAACCCCTCCGAGGAATCTCAACTGCACAAACCCCTACTACACTCCAAATTCAGTAGGA 481
DB 1606 -----CCCCCTGAGGAGACACTACACTGAGGGGCCCATCTTTGGCCCTATCAGAAAGGA 1661

QY 482 AGCAGTTAGAGCAGTGTGTGAGCCAACTCCCAACAGTACTTGGGTTTCTGTTGAGAG 541
DB 1662 AGTAGCTAGAGCAGTCTATTGCCCAA--TTCCCAAGAGCAGCTGGGGTGTCCCGTTTAGAG 1719

QY 542 GGTGGACTGAGAGCAGGACTAGCTGGATTTCTTAGCTGACTAAGATCCCAAGCCTA 601
DB 1720 TGGGGATTGAGAGGTGAAGCCAGCTGACTTCTGGGTGCGGGTGGGAGCTTTGGAGAACTTT 1779

QY 602 NCTGGGAAGGTGACCCGATCCATCTTTTAAACAT 634
DB 1780 TGTGCTAGCTAAAGGATTGTAATGCAACAT 1812

RESULT 15
US-09-374-766-46
; Sequence 46, Application US/09374766
; Patent No. 6579526

GENERAL INFORMATION:
APPLICANT: Herve PERRON
APPLICANT: Frederic BESEME
APPLICANT: Frederic BEDIN
APPLICANT: Glauca PARANHOS-BACCALA
APPLICANT: Florence KOMURIAN-PRADEL
APPLICANT: Colette JOLIVET
APPLICANT: Bernard MANDRAND
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC, THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/374,766
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563
FILING DATE: 02-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 38588
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1859 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-09-374-766-46

Query Match 12.4%; Score 165; DB 4; Length 1859;
Best Local Similarity 58.3%; Pred. No. 1.4e-45;
Matches 369; Conservative 0; Mismatches 247; Indels 17; Gaps 4;

Qy 2 CAAATCGAAGAGCTTTAGACTTGTCTAACCGCCCAAGAGAGGGGGAACCTGTTTATTTT 61
Db 1197 CAAACCGTCAAGCGCTAGACTCTCTACTCTGAGAAAGAGAGACTTGACCTTTCTTA 1256

Qy 62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATCTAGTGAAGAAAGTAAAGAA 121
Db 1257 AGGAGAGAGTGTGTTTACCTAACCCAGTCAGGATAGTA--TGAGATGCTGCCCGGC 1314

Qy 122 ATTTGAGATCAATATATATGATAGCAGAGACCTTCAAACTGACCTGGCGGGGCTC 181
Db 1315 ATTTACAGAAAAGGCTTCTGAAATCAGACACGCTTTTCAAAAT--CTATACCAAC 1370

Qy 182 CTCAGCAATGATGCCCTGGACTCTCCCTTCTTAGGACCTTAGCAGCTATATATTT 241
Db 1371 CTCTGGAGTTGGCAACATGTTTCTTCCCTTCTATGTCCTCCATGGCTGCTTGTCTA 1430

Qy 242 TTACTCCTCTTTGGACCTGTATCTCAACTTCCTTGTAAAGTTTGTCTTTCAGAAAT 301
Db 1431 TTACTCGCTTTGGCGCTGTATTTTAACTCTCTGTCAAAATTTGTTTCTTCTAGGATC 1490

Qy 302 GAGCTGTAAAGCTACAAATAGTTCTTCAATGGAACCCAGATGACGTCCATGACTAAA 361
Db 1491 GAGGCCATCAAGCTACAGATGCTTCAAAATGGAACCCCAATAGCTCAACTATCAAC 1550

Qy 362 ATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGATGTTTAAATGACATTGAA 421
Db 1551 TTCTACTAGGAGCCCTAGAACAAACCCCTGGCCCTTTTCACTGGCCTAAAGAGTT----- 1605

Qy 422 GTCACCCCTCCCGAGAGAAATCTCACTGACAAACCCCTACTACTACCTCCCAATTCAGTAGGA 481
Db 1606 -----CCCTCTCTGGAGGACACTACCACTGCAGGGGCCCCATCTTTTGGCCCTATCCAGAAAGGA 1661

Qy 482 AGCAGTTAGACAGATTGTGACGCAACCTCCCAACAGTACTTTGGGTTTTCCTGTTTGGAGAG 541
Db 1662 AGTAGCTAGAGCAGTCAITTCGCCAA--TTCCCAAGAGAGCTGGGGTGTCCCGTTTAGAG 1719

Qy 542 GGTGACTGAGAGACAGAGACTAGCTGGAATTTCTAGGCTGACTTAAGAAATCCCNAAAGCCTA 601
Db 1720 TGGGGATTGAGAGGTGAAGCCAGCTTGAGCTTCTGGGTGGGTGGGAGCTTGGAGAACTTT 1779

Qy 602 NCTGGGAAGGTGACCGCATCCATCTTTAAACAT 634
Db 1780 TGTGCTAGCTAAAGGATTGTAATGCAACAAT 1812

RESULT 16

US-08-979-847B-42
; Sequence 42, Application US/08979847B
; Patent No. 6582703

GENERAL INFORMATION:

APPLICANT: PERRON, HERVE
BESEME, FREDERIC
BEDIN, FREDERIC
PARANHOS-BACCALA, GLAUCIA
KOMURIAN-PRADEL, FLORENCE
JOLIVET-REYNAUD, COLETTE
MANDRAND, BERNARD
GARSON, JEREMY
TUXE, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC, THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210

CORRESPONDENCE ADDRESS:

ADDRESSEE: OLIFF & BERRIDGE, PLC

STREET: P.O. BOX 19928

CITY: ALEXANDRIA

STATE: VA

COUNTRY: USA

ZIP: 22320

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/979,847B

FILING DATE: 26-No. 6582703-1997

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BERRIDGE, WILLIAM P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: WPB 39046A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 42:

SEQUENCE CHARACTERISTICS:

LENGTH: 1859 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cdna

SEQUENCE DESCRIPTION: SEQ ID NO: 42:

US-08-979-847B-42

Query Match 12.4%; Score 165; DB 4; Length 1859;
Best Local Similarity 58.3%; Pred. No. 1.4e-45;
Matches 369; Conservative 0; Mismatches 247; Indels 17; Gaps 4;

```
QY 2 CAAATCGAAGAGCTTTAGACTTGTCTAACCGCAAAAGAGGGGGAACCTGTTTATTTT 61
DB CAAACCGTCAAGCCCTAGACCTCTCCTGCTGAGAAAGGAGACTCTGACCTTCTTA 1256
QY 62 GGGGAAGAAGTCTTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGAA 121
DB CAAACCGTCAAGCCCTAGACCTCTCCTGCTGAGAAAGGAGACTCTGACCTTCTTA 1256
QY 122 ATTTGAGATCGAATATATATGATAGACGAGGACCTTCAAAACACATGACACCTGGGGCCTC 181
DB AGGAAGAAGTGTGCTCTTACACTTAACCAAGTCAAGGATAGTA--TGAGATGCTGCCCGC 1314
QY 122 ATTTGAGATCGAATATATATGATAGACGAGGACCTTCAAAACACATGACACCTGGGGCCTC 181
DB ATTTGAGAAAGGCTTCTGAATACAGACAAACCCCTTTCAAAATTTT-----CTATACCAAC 1370
QY 182 CTGAGCAATGGAATGATGCTGACCTCTCCCTCTTTAGGACCTCTAGACGCTATAATATTT 241
DB CTCTGGAGTTGGGCAACATGCTTCTTCCCTTTCTATGTCCTGCTGCACTTCTGCTA 1430
QY 242 TTACTCTCTTTGACCTGATCTTCAACTTCTGTTTAAAGTTTGTCTCTTCCAGAAAT 301
DB TTACTCGCTTTGGGCTCTGATTTTAACTCTCTTGTCAAAATTTGTTTCTTCTAGGATC 1490
QY 302 GAGCTCTAAAGCTACAAATAGTTCTTCAATGGAACCCAGATGCACTGACCTATAA 361
DB GAGGCCATCAAGCTACAGATGCTTACAAATGGAACCCCAATGAGCTCAACTATCAAC 1550
QY 362 ATCTACCGTGGACCCCTGGACCGGCTCTGAGACTATGCTCTGATGTTAATGACATTTGAA 421
DB TTCTACTGAGACCCCTAGACCAACCCCTGGCCCTTTCTAGTGCCTAAAGATT-----1605
QY 422 GTCAACCCCTCCGAGGAAATCTCAACTGCAACCCCTACTACACTCCAATTCAGTAGGA 481
DB ----CCCTCTGGAGGACACTACCACTGCGAGGCGCCCTCTTTGCCCCCTATCCAGAAGGA 1661
QY 482 AGCAGTTAGACGAGTGTCTGACCAACCTCCCAACAGTACTTGGGTTTCTGTTGAGAG 541
DB AGTAGCTAGAGCAGTCAATGCCCCAA--TTCCCAAGAGCAGCTGGGGGTCTCCCGTTTAGAG 1719
QY 542 GGTGAGCTGAGACAGGACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCAAGCCTA 601
DB TGGGGATTGAGAGTGAGGAGCAGCTGAGCTTCTGGGTGCGGTGGGAGCTTTGGAGAACTTT 1779
QY 602 NCTGGGAAGGTGACCGCATCATCTTTAAACAT 634
DB TGTGCTAGCTAAAGGATTGTAAATGCAACAAT 1812
```

RESULT 17
US-09-078-294-4
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 80246
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match 12.2%; Score 161.8; DB 3; Length 80246;
Best Local Similarity 59.7%; Pred. No. 2.4e-43;
Matches 360; Conservative 0; Mismatches 234; Indels 19; Gaps 4;

```
QY 2 CAAATCGAAGAGCTTTAGACTTGTCTAACCGCAAAAGAGGGGGAACCTGTTTATTTT 61
DB CAAACCGTCAAGCCCTAGACCTCTCTTACTGCTGAGAAAGGAGATTCTG-CACCTTCTTA 58698
QY 62 GGGGAAGAAGTCTTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGAA 121
DB GGGGTAGAGTGTGTTTATATACTAACCAAGTCAAGGATAATATGAGATACCAACCAAGTGT 58758
QY 122 ATTTGAGATCGAATATATATGATAGACGAGGACCTTCAAAACACATGACACCTGGGGCCTC 181
DB GGGGTAGAGTGTGTTTATATACTAACCAAGTCAAGGATAATATGAGATACCAACCAAGTGT 58758
QY 122 ATTTGAGATCGAATATATATGATAGACGAGGACCTTCAAAACACATGACACCTGGGGCCTC 181
DB TTACAGCAAAAGGCTTCTGAAATCAGACAAATGCTTTCAAACTCT-----TATACCAAC 58812
QY 182 CTGAGCAATGGAATGCTGACCTCTCCCTCTTCTAGGACCTCTAGACGCTATAATATTT 241
DB CTCTGGAGTTGGGCGCATGCTTCTCCCTTTCTAGGCTCTGTGACAGGCACTTCTGCTA 58872
QY 242 TTACTCTCTTTGGACCCCTGTATCTTCAACTTCTTCTTAAAGTTTGTCTCTTCAGAAAT 301
DB ATAGTCGCAATTTGGGCCCTGTATTTTAACTCTTGTGTCAAATTTGTTTCTCTAGGATC 58932
QY 302 GAGCTCTAAAGCTACAAATAGTTCTTCAATGGAACCCAGATGCACTGCACTATAA 361
DB GAGGCCATCAAGCTACAGATGCTTACAAATGTAACCCCAATGAGCTCAACTTAACCAAC 58992
QY 362 ATCTACCGTGGACCCCTGGACCGGCTCTGAGACTATGCTCTGATGTTAATGACATTTGAA 421
DB TTCTGCTGAGGACCCCTGGACCGGCTCTTCT-----AATGGCCTAAAG 59043
QY 422 GTCAACCCCTCCGAGGAAATCTCAACTGCAACCCCTACTACACTCCAATTCAGTAGGA 481
DB AGCTCCCTCTGAGGAGACACTACACTGAGGGGCCCTTCTTACCCCTATCCAGCAGGA 59103
QY 482 AGCAGTTAGACGAGTGTCTAGCCAAACCTCCCAACAGTACTTGGGTTTCTGTTGAGAG 541
DB AGTAGCTACAGCGTCTATGCCAAA--TCCCAACAGCAGCTGGGGGTCTCTGTTTGGAG 59160
QY 542 GGTGAGCTGAGACAGGACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCAAGCCTA 601
DB GGGGGATTGAGAGGTGAAGCCAGCTGGGCTTCTGGGTGAGTGGGGAACCTTT 59220
QY 602 NCTGGGAAGGTGA 614
DB TGTGCTAGCTAA 59233
```

RESULT 18
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA
; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

Query Match 11.4%; Score 151.8; DB 3; Length 80595;
Best Local Similarity 58.5%; Pred. No. 6.5e-40;
Matches 327; Conservative 0; Mismatches 214; Indels 18; Gaps 3;

QY 56 TTTTATAGGGGAAGAATCGTGTAGTATGTTAATCAATCTGGAATCATCTAGGAAAGTT 115
DB TTTTATAGGGGTAGAGTGTGTTTATATACTAAACAGTCAGGATAGTATGAGATACCACC 58985

```
QY 116 AAAGAAATTTGAGATCGAATATATATGTAGACGAGACCTTCAAACACTGCAACCTGG 175
Db 58986 CAGTGTTTACAGGAAAGGGCTTCTGAAATCAGACAATGCGCTTTCAAACCTCT-----TAT 59039
QY 176 GGCTCTCTAGCCAAATGGATGCGCTTGGACTTCCCTTCTTAGGACCTCTAGACGCTATA 235
Db 59040 ACCAACTCTGGAGTTGGGGGACATGCGCTTCTCCCTTTCTAGGTCTGTGACAGCGATC 59099
QY 236 ATATTTTACTCTCTTTGGACCTGTATCTTCAACTTCTCTGTAAAGTTGTCTCTTCC 295
Db 59100 TTGCTAATAGTCGCAATTTGGGCTCTGTATTTTAACTCTTGTCAAATTTGTTCTCT 59159
QY 296 AGAATGAAGCTGTAAAGCTACAAATAGTTCTTCAAATGAACCCAGATGCAATCAATG 355
Db 59160 AGGATCGAGGCCATCAAGCTACAGATGATCTTACAAATGTAACCCCAAAATGAGTCAACT 59219
QY 356 ACTAAATCTACGTTGACCGCTTGGACCGCTGCTAGACTATGCTGTATGTTAATGAC 415
Db 59220 AACAACTTCTGCTGAGGACCGCTTGGACCGCTGCGC-----CTTTCAATGGC 59270
QY 416 ATTGAAGTCAACCTCCCGAGGAAATCTCAACTGCAACACCCCTTACTTACACTCCAATTC 475
Db 59271 CTTAAGAGCTCCCTCTGAGGACACTTACCACCTGCGGGGCCCTTCTTACCCCTATCCA 59330
QY 476 GTAGGAGCAGTTAGACAGTGTGTACGCCAACCTTCCCAACAGTACTTGGGTTTCTGT 535
Db 59331 GCAGGAAGTAGCTACAGCGGTCTATCGCAAA---TCCCAACAGCAGCTGGGTTCTGT 59387
QY 536 TGAGAGGTTGACTGAGACAGGACTAGCTGGATTTCTTAGGCTGACTGAATCCCA 595
Db 59388 TTGAGGGGGGATTTGAGAGGTGAAGCCAGCTGGGCTTCTGGTCAAGTGGGACTTGGAG 59447
QY 596 AGCTTANCTGGGAGGTGA 614
Db 59448 AACTTTGTGTAGCTAA 59466

RESULT 19
US-09-621-976-2599/c
; Sequence 2599, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2599
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 67..228
US-09-621-976-2599

Query Match 9.3%; Score 123.6; DB 4; Length 506;
Best Local Similarity 82.4%; Pred. No. 8.7e-32;
Matches 140; Conservative 10; Mismatches 19; Indels 1; Gaps 1;

QY 923 CATCCACACTGCTGTTGGCCACCGTCAACGCGTGTGACTTCCATCCCTTTGGAT 982
Db 169 CGTCCACCACTGCTGTTGGCCACCGTGTGACTTCCATCCCTTTGGAT 110
QY 983 CCAGCAGAGTCCACTGTGCTCTGATCCAGCAGGTACCATTGGCACTCCCGATCAG 1042
Db 109 CTTGAGGGTGT-CGCTGTGCTCTGATCCAGCAGGGCCCATGGCGCTCCCAATGG 51
QY 1043 GCTAAAGGCTTGCCAATTTCTTCTGATGGCTAAGTGCCTGGGTTTGTCT 1092
```

```
Db 50 GCTAAAGGCTTGCCATTTGTTCTGCGCGCTAAGTKMKKGSYTKGTYT 1

RESULT 20
US-09-120-653D-1
; Sequence 1, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
; APPLICANT: YOON, JI-WON
; APPLICANT: JUN, HEE-SOOK
; APPLICANT: PARK, HAE-JOON
; APPLICANT: PARK, HAE-JOON
; APPLICANT: AHN, JONG-SEONG
; APPLICANT: HA, YOUNG-JU
; APPLICANT: CHUNG, SOO-IL
; TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
; FILE REFERENCE: 98-338
; CURRENT APPLICATION NUMBER: US/09/120,653D
; CURRENT FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: KR 98-10108
; PRIOR FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 1
; LENGTH: 3910
; TYPE: DNA
; ORGANISM: DIABETES-SPECIFIC ENDOGENOUS RETROVIRUS ERV-9
US-09-120-653D-1

Query Match 8.1%; Score 107.8; DB 3; Length 3910;
Best Local Similarity 56.5%; Pred. No. 9.6e-26;
Matches 308; Conservative 0; Mismatches 217; Indels 20; Gaps 5;

QY 2 CAAAATCGAAGAGCTTTAGACTTGTAAACCGCCAAAGAGGGGGAACCTGTTATTTTA 61
Db 2948 CAAAACCGCTGAGGATAGATGTCCTCACTGCTGAGAAGAGAGACTCTGCACCTTCTTA 3007
QY 62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCAATTAAGTAAAGTAAAGAA 121
Db 3008 GGGGAAGAGTGTGTTTGTGACATAACAGGACCGGATAGCA--TGAGATGGCACCCAGC 3065
QY 122 ATTTGAGATCGAATATAATGTAGACGAGGAGCTTCAAACACTGCACCTTGGGCTC 181
Db 3066 GTTTACAGAAAAGGCTGCTGAAAATGAGTGCCTTTTAAATTTCTTATAC-----CAAC 3118
QY 182 CTCAGCCAATGATGCGCTGCGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATATT 241
Db 3119 CTGTGTGTGGGCAACATGGCTTCTCACCTTTCTAGGTCCCGTGCAACCGTCTGGTG 3178
QY 242 TTACTCTCTTTGGACCGCTGTATCTTCAACTTCTTGTAAAGTTTGTCTCTTCAGAAAT 301
Db 3179 TTACTCGCTTTGGGCCAGCAATTTTAAAGTCTTGTCAAAATTTGT-TTGTCTAGATT 3237
QY 302 GAAGCTGTAAAGCTACAAATAGTTCTTCAAATGGAAACCCAGATGCAATGCTGCTATAA 361
Db 3238 GAGGCTTCAAGCTACAGATGGTCTTACATATCAAAACCCCGAATAATTTCAACTACTAC 3297
QY 362 ATCTACCGTGGACCGCTGAGACCGCTGCTAGACTATGCTCTGATGTTAATGACATTGAA 421
Db 3298 TTCTACGGAGACACTTGGACTAACGAGCTGGAGTTACCTTGG-----CCTAGAG 3348
QY 422 GTACCCCTCCGAGGAATCTCAACTGCAACACCCCTACTACATCTCAATTCAGTAGGA 481
Db 3349 AGTTCCCTCTGAAGGTCACCAACTGCAAAAGCCCTTTTGGCGCTATCCATCAGGA 3408
QY 482 AGCAGTTAGCAGTGTGTGAGCCAACTCCCGCAACAGTACTTGGGTTTCTCTGTGAGAG 541
Db 3409 CGTACCTTAGAACAGTCTCTCGGCC-ATATTGCCAACAGCAGTTGAGTGTCTGTGATTG 3467
QY 542 GGTGG 546
Db 3468 AGGGG 3472
```



```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/007.282B
; FILING DATE: 19930121
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR., GERALD M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1644-104P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-007-282B-1

Query Match          6.1%; Score 80.6; DB 1; Length 1704;
Best Local Similarity 53.7%; Pred. No. 1.1e-16;
Matches 167; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 2 CAAATCGAAGAGCTTTAGACTTCTTAACCGCCAAAGAGGGGGAACCTGTTTATTTTA 61
Db 1387 CAAATAGAGAGGGTTAGACCTATTGCTGCGGACCAAGAGGAGGATATGCTCGCACTC 1446
QY 62 GGGGAAGAATGCTGTTAGTATGTTAAATCAATCTGGAATCAATTACTGAGAAAGTTAAAGAA 121
Db 1447 CAGGAGAAGTGTGTTTTTACGCTTAACAAGTCGGGTATCGTGTGACAAGATCCGAAAA 1506
QY 122 ATTTGAGATCGAATATAATGTAGAGCAGAGACCTTCAAAACACTGACCCCTGGGGCCTC 181
Db 1507 CTCGAAGAGGACCTTATCGAGAGAAAAACGTGCACTGTACGACAAACCCCTGTGGAGCGGC 1566
QY 182 CTAGCCAAATGGATGCGCTGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 241
Db 1567 TTGAACGGCTCTCTTCCATATTGCTACCTTGTAGGCCCTGTTGGGCTCATATTG 1626
QY 242 TTACTCTCTTTGACCCCTGATCTCAACTCTGTTGTTAAAGTTGTTCTTCCAGAAAT 301
Db 1627 TTCTGACCTCGCCCGTGCAATTATGAGACCTGACTCGCATTTACATGACAAAAAT 1686
QY 302 GAAGCTGTAAA 312
Db 1687 CAGCAGTAAA 1697

RESULT 24
US-09-011-745-1
; Sequence 1, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weis, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; CURRENT FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2518

; TYPE: DNA
; ORGANISM: RD114
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)
; OTHER INFORMATION: n is any nucleotide
US-09-011-745-1

Query Match          6.1%; Score 80.6; DB 3; Length 2518;
Best Local Similarity 53.7%; Pred. No. 1.5e-16;
Matches 167; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 2 CAAAATCGAAGAGCTTTAGACTTGTCTAACCAGCAAGAGGGGGAACCTGTTTATTTTA 61
Db 1930 CAAAATAGGAGGGGACTGGACCTACTAACCAGCAGAACAGGAGGAATTTGTTTAGCCTTA 1989
QY 62 GGGGAAGAATGCTGTTAGTATGTTAAATCAATCTGGAATCAATTACTGAGAAAGTTAAAGAA 121
Db 1990 CAAGAAAAATGCTGTTTATGCTAAACAAGTCAGGAATTTGTGAGAAACAAAAATAAGAAC 2049
QY 122 ATTTGAGATCGAATATAATGTAGAGCAGAGACCTTCAAAACACTGACCCCTGGGGCCTC 181
Db 2050 CTACAAGAAGATTACAAAACGACGGGAAAGCCTGGCAACCAACCTCTCTGGACCGGG 2109
QY 182 CTAGCCAAATGGATGCGCTGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 241
Db 2110 CTGACGGGCTTTCTCCGTACCTCTTACCTCTCTCTGGGACCCCTACTACCCCTCTCTAC 2169
QY 242 TTACTCTCTTTGACCCCTGATCTCAACTCTGTTGTTAAAGTTGTTCTTCCAGAAAT 301
Db 2170 ATACTAACCATTTGGGCGCATCGCTTTTTCAGTCGCTCATGCGCTTCATTAATGATAGACT 2229
QY 302 GAAGCTGTAAA 312
Db 2230 AATGTTGTACA 2240

RESULT 25
US-09-555-352-1
; Sequence 1, Application US/09555352
; Patent No. 6544779
; GENERAL INFORMATION:
; APPLICANT: Cichutek, Klaus
; APPLICANT: Mergert-Millitzer, Heike
; TITLE OF INVENTION: PSUDO-TYPE RETROVIRAL VECTORS WITH
; TITLE OF INVENTION: MODIFIABLE SURFACE CAPSID PROTEINS
; FILE REFERENCE: 11692-005001
; CURRENT APPLICATION NUMBER: US/09/555,352
; CURRENT FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: PCT/DE98/03542
; PRIOR FILING DATE: 1998-11-27
; PRIOR APPLICATION NUMBER: DE 197 52 855.4
; PRIOR FILING DATE: 1997-11-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4776
; TYPE: DNA
; ORGANISM: Murine leukemia virus
US-09-555-352-1

Query Match          6.1%; Score 80.6; DB 4; Length 4776;
Best Local Similarity 53.7%; Pred. No. 2.4e-16;
Matches 167; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 2 CAAAATCGAAGAGCTTTAGACTTGTCTAACCAGCAAGAGGGGGAACCTGTTTATTTTA 61
Db 1406 CAAAATAGAGAGGGTTAGACCTATTGACTGCCGCAACAGGAGGAATATGTCGCACTC 1465
QY 62 GGGGAAGAATGCTGTTAGTATGTTAAATCAATCTGGAATCAATTACTGAGAAAGTTAAAGAA 121
Db 1466 CAGGAGAAGTGTGTTTTTTTACGCTAACCAAGTCGGGTATCGTACGTGACAGATCCGAAA 1525
```

```
QY 122 ATTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCTGGGGCTC 181
Db 1526 CTCGAGAGGACCTTATCGAGAGAAACGTGCACTGTACACACCCCTGTGGAGGGC 1585

QY 182 CTCAGCCAATGGATGCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATATTT 241
Db 1586 TTGAACGGCTTCTTCCATATTTGCTACCTTGTAGGCCCTCTGTGGCTCATATTG 1645

QY 242 TTACTCCTCTTTGACCGCTGTATCTCAACTTCTTGTGTTAAGTTTGTCTTCCAGAAAT 301
Db 1646 TTCTGACCTCGGCCGTGCATTATGAGACCGCTGACTGCGATTATACATGACAAAAT 1705

QY 302 GAAGCTGTAAA 312
Db 1706 CAGCAGTAAA 1716

RESULT 26
US-09-011-745-8
; Sequence 8, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi Yasuhiro
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; CURRENT FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 5865
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; OTHER INFORMATION: Construct
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3611)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3612)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3613)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3614)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3799)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3800)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3801)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (3802)
; OTHER INFORMATION: n is any nucleotide
US-09-011-745-8

Query Match 6.1%; Score 80.6; DB 3; Length 5865;
Best Local Similarity 53.7%; Pred. No. 2.7e-16;
Matches 167; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 2 CAAAATCGAAGAGCTTTAGACTTGTAAACGCCAAAGAGGGGAAACCTGTTTATTTTAA 61
Db 2788 CAAAATAGAGGGGAGCTGGACCTACTAAACGGCAGAACAAAGAGGAATTTGTTAGCCTTA 2847

QY 62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGATCATTTACTGAGAAAGTTAAGAA 121
Db 2848 CAAGAAAAATGCTGTTTTTTTATGCTAAACAAGTCAGGAATTTGTGAGAACAAAAAAGAACC 2907

QY 122 ATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCTCGGGGCTC 181
Db 2908 CTACAAGAAGATTACAAAACGCAGGGAAGCTGGCAACCAACCTCTCTGACCGGG 2967

QY 182 CTCAGCCAATGGATGCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATTT 241
Db 2968 CTGACGGGCTTCTTCCGTACCTCTACCTCTCTGGGACCCCTACTCACCTCTCTACTC 3027

QY 242 TTACTCCTCTTTGGACCGCTGTATCTTCAACTTCTTGTAAAGTTTGTCTTCCAGAAAT 301
Db 3028 ATACTAACCATTTGGGCCATGCGTTTTTCAGTCGCGCTCATGGCCTTCATTATGATAGACTT 3087

QY 302 GAAGCTGTAAA 312
Db 3088 AATGTTGTACA 3098

RESULT 27
US-09-120-653D-6
; Sequence 6, Application US/09120653D
; Patent No. 6385727
; GENERAL INFORMATION:
; APPLICANT: YOON, JI-WON
; APPLICANT: JUN, HEE-SOOK
; APPLICANT: PARK, HAE-JOON
; APPLICANT: AHN, JONG-SEONG
; APPLICANT: HA, YOUNG-JU
; APPLICANT: CHUNG, SOO-IL
; TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
; FILE REFERENCE: 98-338
; CURRENT APPLICATION NUMBER: US/09/120,653D
; CURRENT FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: KR 98-10108
; PRIOR FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 6
; LENGTH: 464
; TYPE: DNA
; ORGANISM: DIABETES-SPECIFIC RETROVIRUS ENDOGENOUS ERV-9
US-09-120-653D-6

Query Match 5.9%; Score 78.2; DB 3; Length 464;
Best Local Similarity 55.9%; Pred. No. 3.1e-16;
Matches 213; Conservative 0; Mismatches 158; Indels 10; Gaps 3;

QY 2 CAAAATCGAAGAGCTTTAGACTTGTAAACGCCAAAGAGGGGAAACCTGTTTATTTTAA 61
Db 91 CAAAACCGCTGAGGCATAGATGCTCTACTGCTGAAGAAAGAGGACTCTGCACCTTCTTA 150

QY 62 GGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAAATCATTTACTGAGAAAGTTAAGAA 121
Db 151 GGGGAAGAGTGTGTTTTCACATTAACAGGACCGATAGCA--TGAGATGGCACCAGC 208

QY 122 ATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTCAAAACACTGCACCTCGGGGCTC 181
Db 209 GTTTACAGAAAAAGGCTGCTGAATAGTCGCGCTTTTAAATTTCTTATAC-----CAAC 261
```



```
QY 1 TCAAAATCGAAGAGCTTTAGACTTGCTAACGCCAAGAGAGGGGACCTGTTATTTTT 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1995 TCAAAACTGGAGATCCTGGACCTCCTAACCCACTGAGAAAGGGGTACCTGCATATACCT 2054
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 AGGGGAAGAATGCTGTGTATGTTAATCAATCTGGAATCAATTACTGAGAAAGTTAAAGA 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2055 CCAGGAAGAATGCTGTGTGTGTAAATGAATCTGGCATTTTCATATCGCAGTTGCTAG 2114
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 AATTGAGATCG---AATATAATGTAGACGAGGACCTTCAAAACACTGCACCCCTGGGG 177
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2115 GCTTCATGACAGGGCTGCAGAGCTTAGACATCAAGTCGCTGACTCCTGTGGCAAGGATC 2174
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 178 CCTCCTCAGCCAATGGATGCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAAT 237
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2175 ATCCCTCTAAGATGGATACCTGGTTGCCCTTCTTAGGACCCCTGATCTTCTCTT 2234
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 238 ATTTTACTCCTCTTTGGACCCCTGTATCTTCAACTT 273
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2235 CCTGTTACTAATGATTGGGCCCATGCATATTTAACTT 2270
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: January 22, 2005, 19:37:03
Job time : 127.063 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2005, 12:12:40 ; Search time 721.788 Seconds
(without alignments)
10579.682 Million cell updates/sec

Title: US-09-319-156B-12
Perfect score: 1329
Sequence: 1 tcaaatcagagcttag.....gcaagatccccagtaaca 1329

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1326	99.8	1329	8	US-08-979-847-108
2	1326	99.8	1329	15	US-10-114-104-108
3	1035.2	77.9	56093	10	US-09-873-367C-81
4	798.2	60.1	21646	13	US-10-087-192-910
5	778.8	58.6	1393	16	US-10-220-120-17
6	775.4	58.3	2030	14	US-10-637-565-18
7	774.2	58.3	7974	14	US-10-198-846-9936
8	766.8	57.7	2074	16	US-10-416-642-4
9	731.8	55.1	2946	13	US-10-114-893-134
10	731.8	55.1	2946	15	US-10-016-249-3
11	730.8	55.0	2930	9	US-09-902-535-1
12	722.6	54.4	7582	16	US-10-632-793-30

13	721.2	54.3	1136	16	US-10-632-793-25	Sequence 25, Appl	
14	712.6	53.6	2782	16	US-10-632-793-26	Sequence 26, Appl	
15	711	53.5	2782	16	US-10-133-036-1	Sequence 1, Appl	
16	704.8	53.0	1894	9	US-09-864-761-4444	Sequence 4444, App	
C	17	643.4	48.4	161334	13	US-10-087-192-730	Sequence 730, App
C	18	637.8	48.0	780	10	US-09-854-867-385	Sequence 385, Appl
C	19	601.4	45.3	157090	17	US-10-672-766A-34	Sequence 34, Appl
C	20	579.8	43.6	180557	13	US-10-003-806-6	Sequence 6, Appli
C	21	579.8	43.6	180557	13	US-10-003-808-9	Sequence 9, Appli
22	570.2	42.9	161671	15	US-10-017-117-1	Sequence 1, Appli	
23	545.4	41.0	134292	16	US-10-240-425-1102	Sequence 1102, Ap	
24	545	41.0	635	8	US-08-979-847-102	Sequence 102, App	
25	545	41.0	635	15	US-10-114-104-102	Sequence 102, App	
26	536.2	40.3	849	16	US-10-220-120-15	Sequence 15, Appl	
C	27	530.6	39.9	366710	18	US-10-719-993-6792	Sequence 6792, Ap
C	28	512	38.5	17758	16	US-10-264-237-2833	Sequence 2833, Ap
C	29	502	37.8	326014	9	US-09-731-231A-3	Sequence 3, Appli
C	30	502	37.8	326014	17	US-10-751-985-3	Sequence 3, Appli
C	31	478.6	36.0	283351	18	US-10-719-993-7065	Sequence 7065, Ap
32	478.6	36.0	1980090	18	US-10-719-993-6815	Sequence 6815, Ap	
33	468.8	35.3	2052	16	US-10-276-774-678	Sequence 678, App	
34	465	35.0	3372	16	US-10-632-793-28	Sequence 28, Appl	
35	465	35.0	22436	11	US-09-997-722-148	Sequence 148, App	
36	460	34.6	285020	13	US-10-087-192-1666	Sequence 1666, App	
C	37	457.8	34.4	2145	13	US-10-027-632-98772	Sequence 98772, A
C	38	457.8	34.4	2145	15	US-10-027-632-98772	Sequence 98772, A
39	455.8	34.3	1003	16	US-10-637-565-1	Sequence 1, Appli	
40	455.8	34.3	2052	16	US-10-637-565-14	Sequence 14, Appl	
C	41	453.6	34.1	559	9	US-09-864-761-7501	Sequence 7501, Ap
42	444	33.4	581	15	US-10-029-386-5948	Sequence 5948, Ap	
43	440.4	33.1	8523	10	US-09-854-867-21	Sequence 21, Appl	
44	440.2	33.1	1321	16	US-10-632-793-20	Sequence 20, Appl	
45	438	33.0	2349	15	US-10-104-047-1148	Sequence 1148, Ap	

ALIGNMENTS

RESULT 1

US-08-979-847-108
; Sequence 108, Application US/08979847
; Publication No. US20030039664A1
; GENERAL INFORMATION:
; APPLICANT: FERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
; APPLICANT: TUKE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESS: OLIFF & BERRIDGE, PLC
; STREET: P. O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,847
; FILING DATE: 26-NOV-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

; NAME: BERRIDGE, WILLIAM P. ; REGISTRATION NUMBER: 30,024 ; REFERENCE/DOCKET NUMBER: WPB 39046A ; TELECOMMUNICATION INFORMATION: ; TELEPHONE: 703-836-6400 ; TELEFAX: 703-836-2787 ; INFORMATION FOR SEQ ID NO: 108: ; SEQUENCE CHARACTERISTICS: ; LENGTH: 1329 base pairs ; TYPE: nucleic acid ; STRANDEDNESS: single ; TOPOLOGY: linear ; MOLECULE TYPE: cdna ; US-08-979-847-108									
Query Match 99.8%; Score 1326; DB 8; Length 1329; Best Local Similarity 100.0%; Pred. No. 0; Matches 1329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	TCAAAATCGAAGAGCTTTAGACTTGCTAAACCGCAAAAGAGGGGGAACCTGTTTATTTT	60						
DB	1	TCAAAATCGAAGAGCTTTAGACTTGCTAAACCGCAAAAGAGGGGGAACCTGTTTATTTT	60						
QY	61	AGGGGAAGAATGCTGTAGTATGTTAATCAATCTGGAATCATTAATCTGAGAAAGTTAAAGA	120						
DB	61	AGGGGAAGAATGCTGTAGTATGTTAATCAATCTGGAATCATTAATCTGAGAAAGTTAAAGA	120						
QY	121	AATTTGAGATCGAATAATATGTAGACAGAGGACCTTCRAAACACACTGCACCCCTGGGCGCT	180						
DB	121	AATTTGAGATCGAATAATATGTAGACAGAGGACCTTCRAAACACACTGCACCCCTGGGCGCT	180						
QY	181	CCTCAGCAATGATGCGCTGGACTCTCCCTCTTTAGGACCTCTAGCAGCTATAATATT	240						
DB	181	CCTCAGCAATGATGCGCTGGACTCTCCCTCTTTAGGACCTCTAGCAGCTATAATATT	240						
QY	241	TTTACTCCTCTTTGGACCCCTGTAATCTTCAACTTCTTGTAAAGTTTGTCTCTCCAGAA	300						
DB	241	TTTACTCCTCTTTGGACCCCTGTAATCTTCAACTTCTTGTAAAGTTTGTCTCTCCAGAA	300						
QY	301	TGAAGCTGTAAAGCTCAAAATAGTTCTTCAAAATGGAACCCAGATCGATCCATGACTAA	360						
DB	301	TGAAGCTGTAAAGCTCAAAATAGTTCTTCAAAATGGAACCCAGATCGATCCATGACTAA	360						
QY	361	AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGTTAAATCACAATTGA	420						
DB	361	AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGTTAAATCACAATTGA	420						
QY	421	AGTCACCCCTCCGAGGAATCTCAACTGCAACACCCCTACTACACTCCTCAATTCAGTAGG	480						
DB	421	AGTCACCCCTCCGAGGAATCTCAACTGCAACACCCCTACTACACTCCTCAATTCAGTAGG	480						
QY	481	AAGCAGTTAGACAGTGTGTCAGCAACCCCTCCCAACAGTACTTGGGTTTCCGTGTGAGA	540						
DB	481	AAGCAGTTAGACAGTGTGTCAGCAACCCCTCCCAACAGTACTTGGGTTTCCGTGTGAGA	540						
QY	541	GGGTGACTCAGACAGGACTAGCTGGATTCTCTAGGCTGACTAAGAAATCCCNAAAGCCT	600						
DB	541	GGGTGACTCAGACAGGACTAGCTGGATTCTCTAGGCTGACTAAGAAATCCCNAAAGCCT	600						
QY	601	ANTCTGGAAGGTACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG	660						
DB	601	ANTCTGGAAGGTACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG	660						
QY	661	ACCAATCAGAGCTCACAATAATGCTAATCAGGCAAAACAGAGGTAAGCAATAGCC	720						
DB	661	ACCAATCAGAGCTCACAATAATGCTAATCAGGCAAAACAGAGGTAAGCAATAGCC	720						
QY	721	AATCATCTATTGCTGAGACACAGCGGGGAAGACAGGATTCGGGATATAAACTCAGGCA	780						
DB	721	AATCATCTATTGCTGAGACACAGCGGGGAAGACAGGATTCGGGATATAAACTCAGGCA	780						
QY	781	TTCAAGCCAGCAACAGAACCCCTTTGGGTCCCTCCCATTTGATGGAGCTCTCTTTT	840						

RESULT 2
US-10-114-104-108
; Sequence 108, Application US/10114104
; Publication No. US20030198647A1
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; BESEME, FREDERIC
; BEDIN, FREDERIC
; PARANHOS-BACCALA, GLAUCIA
; KOMURIAN-PRADEL, FLORENCE
; JOLIVET-REYNAUD, COLETTE
; MANDRAND, BERNARD
; GABSON, JEREMY
; TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/10/114,104
APPLICATION NUMBER: US/10/114,104
FILING DATE: 03-Apr-2002
CLASSIFICATION: <Unknown>

[illegible]


```
Db 240 CGGGGACTGAGACAGAGACTAGCTGGATTCTTAGGCTGACTAAGAAATCCCTTAAGCCTTA 299
QY 602 NCTGGGAAGTGACCCGATCCATCTTTAAACATGGGGCTTGCAACTTAGCTACACCCGA 661
Db 300 GCTGGGAAGTGACCCGCTCTACCTTTAAACAGGGGCTTGCAACTTAGCTACACCCGA 359
QY 662 CCAATC-----AGAGAGCTCACTAAATCTAATCAGGCAAAACAGGAGGTAAAG 712
Db 360 CCAATCAGGTAGTAAAGAGAGCTCACTAAATCTAATAGGCAAAACAGGAGGTAGAG 419
QY 713 CAATAGCAATCATCTATTGCTGAGAGCAACAGGGGAAGGACAAGATGGGATATAA 772
Db 420 AATATGCCAATCATCTATCGCTGAGAGCAACAGGAGGAGCAATGATCCGGATATAA 479
QY 773 CTGAGGCAATCAAGCAGAGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGATGGAGC 832
Db 480 CCAAGCATTCGAGCAGCAACAGGCTGCCCCCTTTGTGTCCTCTCTTTGATGGAGC 539
QY 833 TCTGTTTTCACTCTATTTCACCTCTATTAAATCAATGCAACTGCACTCTTCTGTFCCGTGT 892
Db 540 TCTGT-----TTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGTFCTGTGT 589
QY 893 TTTTATGGCTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTGCCACCGTCACA 952
Db 590 TGTACGGTTTGAGCTGAGCTTTTCGCTCGCGCTCCACCACTGCTGTTGCCGCCATCGGA 649
QY 953 GACCCGCTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCACCTGTGCTCCTGATCC 1012
Db 650 GACCTGCGCTGACTTCCATCCCTCCGGATCTGGCAGGGTGTTCA-TGTGCTCCTGATCC 708
QY 1013 AGCGAGGTACCCATTCGCCATCCCGATCAGGCTAAAGGGTTGCCATTTGTTCTGCAATGC 1072
Db 709 AGAGAGGACCCCATTCGCATTCCTGATGGGCTAAAGGCTTGCCATTTGTTCTGCAAGAC 768
QY 1073 TAAGTGCCTGGGTTTGTCTTAATAGAACTGAACTGCTGCTGCTGCTGCTGCTGCTGCT 1132
Db 769 TAAGTGCCTGGGTTTGTCTTAATAGAACTGAACTGCTGCTGCTGCTGCTGCTGCTGCT 828
QY 1133 TCCATGACCCACAGGCTTCTTAATAGAGCTATAACACTCACCGCATGCCCAAGATTCAT 1192
Db 829 TCCGTGACCCACAGCTTCTTAATAGAGTATAACACTCACCGCAGGCCCAAGATTCAT 888
QY 1193 CTTTGGTATCTGAGGCCAAGAACCCAGGTGAGAGAGTGTGAGCTTGCCACCATTTG 1252
Db 889 CTTTGAATCCGTTAGGSCCAAGAACCCAGGTGAGAGATGAGGCTTGCCACCATTTT 948
QY 1253 GGAAGTGGCCCATGCGCATTTGGTAGCGGCCCAACCACTCTTGGAGCTGTGGAGCA 1312
Db 949 GGAAGTGGCTGCGGCCCATTTGGAGTGGCTTGCCACCATCTTGGAGCTGTGGAGCA 1008
QY 1313 AGGATCCCCCAGTAACA 1329
Db 1009 AGGATCCCCCTGGTAACA 1025
```

RESULT 6

```
US-10-637-565-18
; Sequence 18, Application US/10637565
; Publication No. US20040043381A1
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucaia
; APPLICANT: PERRON, Hervé
; APPLICANT: KOMURIAN-PRADEL, Florence
; TITLE OF INVENTION: THE LTR REGION OF MSRV-1 AND THE PROTEINS IT ENCODES, AND PROBES
; FILE OF INVENTION: METHODS FOR DETECTING MSRV-1 RETROVIRUS
; FILE REFERENCE: 110257
; CURRENT FILING DATE: 2003-08-11
; PRIOR FILING DATE: US/09/890,340
; PRIOR FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: PCT/IB00/00159
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: EP 99420041.8
```

```
; PRIOR FILING DATE: 1999-02-15
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 2030
; TYPE: DNA
; ORGANISM: MSRV-1 retrovirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1626)
US-10-637-565-18
```

Query Match 58.3%; Score 775.4; DB 16; Length 2030;

Best Local Similarity 92.8%; Pred. No. 1.9e-230;

Matches 812; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
QY 1 TCAAAATCAAGAGCTTTAGACTTGCTAACCCGCAAAAGAGGGGAACCTGTTTATTTTT 60
Db 1140 TCAAAATCAAGAGCTTTAGACTTGCTAACCCGCAAAAGAGGGGAACCTGTTTATTTTT 1199
QY 61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTTCTGAGAAAGTTAAAGA 120
Db 1200 AGGAGAGAACGCTGTTATGTTAATCAATCCAGAAATTGTCACGTGAGAAAGTTAAAGA 1259
QY 121 AATTGAGATCGAATATAATGTAGACGAGAGACCTTCAAAACACATGCAACCTGGGGCCT 180
Db 1260 AATTCAGATCGAATACAAATGTAGACGAGAGGCTTCAAAACACACGACGCTGGGGCCT 1319
QY 181 CCTCAGCCAATCGATGCCCTGGACTCTCCCTCTTCTAGGACCTCTAGCAGCTATAATTT 240
Db 1320 CCTCAGCCAATCGATGCCCTGGGTTCTCCCTCTTCTAGGACCTCTAGCAGCTATAATTT 1379
QY 241 TTTACTCTCTTTGGACCCCTGTATCTTCAACTTCTTGTGTTAAAGTTTGTCTTCCAGAAT 300
Db 1380 GTTACTCTCTTTGGACCCCTGTATCTTTAACTCTTCTGTTAAAGTTTGTCTTCCAGAAT 1439
QY 301 TGAAGCTGTAAAGCTACAAATAGTCTTCAATGGAACCCAGATGCACTGATGACTAA 360
Db 1440 TGAAGCTGTAAAGCTACAGATGCTCTTCAAAATGGAACCCAGATGCACTGATGACTAA 1499
QY 361 AATCTACCCCTGGACCCCTGGACCGGCTCTGCTAGACTATCTCTGATGTTAACTGACTGA 420
Db 1500 GATCCACCCCTGGACCCCTGGACCGGCTCTGCTAGCCCATGCTCGATGTTAACTGACTGA 1559
QY 421 AGTCACCCCTCCGAGGAAATCTCACTGCACAAACCCCTACTACACTCCAATTCAGTAGG 480
Db 1560 AGGCACCCCTCCGAGGAAATCTCACTGCACAAACCCCTACTATGCCCAATTCAGCGGG 1619
QY 481 AAGCAGTTAGAGAGTTGTGAGCAACCTCCCAACAGTACTTTGGGTTTCTGTTGAGA 540
Db 1620 AAGCAGTTAGAGCGGTTCATCAGCCAACTCCCAACAGCAGCTTTGGGTTTCTGTTGAGA 1679
QY 541 GGGTGAGCTGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCCAAGACCT 600
Db 1680 GGGGGAGCTGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCCTAAGCCT 1739
QY 601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
Db 1740 AGCTGGGAAGGTGACTGCATCCACCTCTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 1799
QY 661 ACCAATCAGAGAGCTCACTAAATGTATTCAGGCAAAACAGGAGGTAAAGCAATAGCC 720
Db 1800 ACCAATCAGAGAGCTCACTAAATGTATTCAGGCAAAACAGGAGGTAAAGCAATAGCC 1859
QY 721 AATCATCTATTCCTGAGAGCAACAGCGGGAAGGACAGGATTCGGGATATAAACTCAGGCA 780
Db 1860 AATCATCTATTCCTGAGAGCAACAGCGGGAAGGACAGGATTCGGGATATAAACTCAGGCA 1919
QY 781 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCAATGTTGATGGAGCTCTGTTTT 840
Db 1920 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCTTTGTTGATGGGGCTCTGTTTT 1979
QY 841 CACTCTATTTCACCTCTATTAAATCATGCAACTGCA 875
```



```
Db      2835  TTGAGCGGCAAGCGCAACCCCTTTGGTCCCTCCCTTTGTATGGAGCTCTGTTTT 2894
QY      841  CACTCTATTCTCTATTAAATCATGCAACTGCA 875
Db      2895  CATGCTATTCTCTATTAAATCTTGCAACTGCA 2929

RESULT 10
US-10-016-249-3
; Sequence 3, Application US/10016249
; Publication No. US20030100053A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Mi, Sha
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 6006B.AJ172A
; CURRENT APPLICATION NUMBER: US/10/016,249
; PRIOR FILING DATE: 2001-10-30
; PRIOR FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2946
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-016-249-3

Query Match      55.1%; Score 731.8; DB 15; Length 2946;
Best Local Similarity 90.7%; Pred. No. 9.1e-217;
Matches 794; Conservative 0; Mismatches 69; Indels 12; Gaps 1;

QY      1  TCAAAATCGAAGAGCTTACACTTGCTTAACCGCCAAAGAGGGGGAACCTGTTATTTT 60
Db      2067  TCAAAATCGAAGAGCTTACACTTGCTTAACCGCTGAAGAGGGGGAACCTGTTATTTT 2126
QY      61  AGGGGAAGAAATGCTGTTAGTATGTTAATCAATCGGAATCATCTGCAAGAAAGTTAAAGA 120
Db      2127  AGGGGAAGAAATGCTGTTATGTTAATCAATCGGAATCGTCACTGAGAAAGTTAAAGA 2186
QY      121  AATTGAGATCGAATATAATGTATAGAGAGAGACCTTCAAAACACTGCACTGGGGCCT 180
Db      2187  AATTGAGATCGAATATAACAGTATAGAGAGAGAGCTTCGAAACACTGGAACCTGGGGCCT 2246
QY      181  CCTCAGCAATGATGCCCTGGAGCTCTCCCTTCTTAGGACCTCTAGCAGCTATATATT 240
Db      2247  CCTCAGCAATGATGCCCTGGAGCTCTCCCTTCTTAGGACCTCTAGCAGCTATATATT 2306
QY      241  TTCTACTCTCTTTGGACCTCTGATCTTCAACTTCTTGTAGTTGTTCTTCCAGAAAT 300
Db      2307  GCTACTCTCTTTGGACCTCTGATCTTAACTCTCTTGTAACTTGTCTCTCCAGAAAT 2366
QY      301  TGAAGCTGTAAGACTACAAATAGTTCTTCAAAATGGAACCCAGATGCACTGCAATGCTAA 360
Db      2367  CGAAGCTGTAAACTA-----CAAAATGGAGCCCAAGATGCACTGCAAGACTAA 2414
QY      361  AATCTACGTTGACCCCTGGACCGGCTGTAGACTATGCTGTGATGTTAAATGACATTA 420
Db      2415  GATCTACGCGAGACCCCTGGACCGGCTGTAGCCCAAGCTCTGATGTTAAATGACATTA 2474
QY      421  AGTCACCCCTCCGAGGAAATCTCAATGCAACACCCCTACTACACTCCAAATTCAGTAGG 480
Db      2475  AGGCACCCCTCCGAGGAAATCTCAGTGCACAACTCTACTACGCCCAATTCACAGG 2534
QY      481  AAGCAGTTAGAGCAGTTGTGTCAGCCAACTCCCAACAGTACTTGGGTTTTCTGTTGAGA 540
```

```
Db      2535  AAGCAGTTAGAGCGGTGCTGGCCAACCTCCCAACAGACACTTAGGTTTCTGTTGAGA 2594
QY      541  GGGTGGAGTGAGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAAGAATCCCNAAAGCCT 600
Db      2595  TGGGGGACTGAGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAAGAATCCCTAAGCCT 2654
QY      601  ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCG 660
Db      2655  AGCTGGGAAGGTGACCAATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2714
QY      661  ACCAATCAGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGTAAAGCAATAGCC 720
Db      2715  ACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGTAAAGCAATAGCC 2774
QY      721  AATCATCTATTGCTGAGAGACAGCGGGAAGGACAAAGGATTGGGATATATAAATCAGGCA 780
Db      2775  AATCATCTATTGCTGAGAGACAGCGGAGGAGGACAAATGATCGGATATATAAATCAGGCA 2834
QY      781  TTCAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCAATGTTATGGGAGCTCTGTTTT 840
Db      2835  TTCAGCCGCGCAACGCGCAACCCCTTTGGGTCCCTCCCTTTGTTATGGGAGCTCTGTTTT 2894
QY      841  CACTCTATTCTCACTCTATTAAATCATGCAACTGCA 875
Db      2895  CATGCTATTCTCACTCTATTAAATCTTGCAACTGCA 2929

RESULT 11
US-09-902-535-1
; Sequence 1, Application US/09902535
; Patent No. US20020102530A1
; GENERAL INFORMATION:
; APPLICANT: Keith, Jr., James C.
; APPLICANT: McCoy, John M.
; APPLICANT: Mi, Sha
; TITLE OF INVENTION: Methods and compositions for diagnosing
; TITLE OF INVENTION: and treating preclampsia and gestational trophoblast
; FILE REFERENCE: GIN-6006B4
; CURRENT APPLICATION NUMBER: US/09/902,535
; CURRENT FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 60/216,657
; PRIOR FILING DATE: 2000-07-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2930
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (930)...(2546)
US-09-902-535-1

Query Match      55.0%; Score 730.8; DB 9; Length 2930;
Best Local Similarity 90.7%; Pred. No. 1.9e-216;
Matches 793; Conservative 0; Mismatches 69; Indels 12; Gaps 1;

QY      1  TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAGAGGGGGAACCTGTTATTTT 60
Db      2069  TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAGAGGGGGAACCTGTTATTTT 2128
QY      61  AGGGGAAGAAATGCTGTTAGTATGTTAATCAATCTGGAATCATCTAGCAAGCTTAAAGA 120
Db      2129  AGGGGAAGAAATGCTGTTATTTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2188
QY      121  AATTTGAGATCGAATATAATGTAGAGCAGAGGACCTTTCAAAACACTGCACTGGGGCCT 180
Db      2189  AATTCGAGATCGAATATAACAGTATAGACAGAGAGCTTCGAAACACTGGACCTGGGGCCT 2248
QY      181  CTTAGCCCAATGGATGCCCTTGAGCTCTCCCTTTCTTAGGACCTCTTAGCAGCTATATATT 240
```

```
Db 2249 CCTCAGCCAAATGATGCGCTGGATTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATATT 2308
QY 241 TTACTCCTCTTTGGACCCCTGTATCTTCAACTTCCCTTTGTTAACTTTGTCTCTCCAGAAAT 300
Db 2309 GCTACTCCTCTTTGGACCCCTGTATCTTAACTTCCCTTTGTTAACTTTGTCTCTCCAGAAAT 2368
QY 301 TGAAGCTGTAAAGCTTCAAAATAGTTCTTCAAATGGAAACCCAGATGAGTCCATGACTAA 360
Db 2369 CGAAGCTGTAAACTA-----CAAATGGAGCCCAAGATGAGTCCCAAGACTAA 2416
QY 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGATGTTAAAGACATTGA 420
Db 2417 GATCTACCGCAGACCCCTGGACCGGCTGTAGCCCAAGCTCTGATGTTAAAGACACTAA 2476
QY 421 AGTCACCCCTCCGAGGAAATCTCAACTGACAAACCCCTACTACACTCCAAATTCAGTAGG 480
Db 2477 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACTCTACTACGCCCAATTCAGCAGG 2536
QY 481 AAGCAGTTAGAGCAGTTGTGAGCCAAACCTCCCAAGCAGTACTTTGGGTTTCTCTGTTGAGA 540
Db 2537 AAGCAGTTAGAGCGGTGCTCGGCCAAACCTCCCAAGCAGTACTTTGGGTTTCTCTGTTGAGA 2596
QY 541 GGTGAGCTGAGAGCAGACTAGCTGGATTTCCTAGGCTGACTAAGAAATCCCAAGCCT 600
Db 2597 TGGGGGACTGAGAGCAGACTAGCTGGATTTCCTAGGCTGACTAAGAAATCCCAAGCCT 2656
QY 601 ANCTGGGAAGGTGACCGCTCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACAACCCG 660
Db 2657 AGCTGGGAAGGTGACCACTCCACCTTTAAACAGCGGGCTTGCAACTTAGCTCACAACCTG 2716
QY 661 ACCAATCAGAGCTCCTCAATAATGCTAATCAGGCAAAACAGGAGTAAAGCAATAGCC 720
Db 2717 ACCAATCAGAGCTCCTCAATAATGCTAATCAGGCAAAACAGGAGTAAAGCAATAGCC 2776
QY 721 AATCATCTATTGCTGAGAGCAGCGGAGAGCAAGATTCGGATATAACTCAGGCA 780
Db 2777 AATCATCTATTGCTGAGAGCAGCAGGAGGAGCAATGATCGGATATAAACCCAAGTC 2836
QY 781 TTCAAGCAGCAACAGCAACCCCTTTGGTCCCTCCCACTTATGAGGAGCTCTGTTT 840
Db 2837 TTGAGCGGCAACGCAACCCCTTTGGTCCCTCCCTTTGATGAGGAGCTCTGTTT 2896
QY 841 CACTCTATTTCATCTTATTAATCATGCAACTGC 874
Db 2897 CATGCTATTTCATCTTATTAATCTTGCAACTGC 2930

RESULT 12
US-10-632-793-30
; Sequence 30, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: MALLET, Francois
; APPLICANT: VOISSET, Cecile
; TITLE OF INVENTION: ENDOGENOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
; TITLE OF INVENTION: AUTOIMMUNE DISEASE, LABELING METHOD AND REAGENT
; FILE REFERENCE: 110048
; CURRENT APPLICATION NUMBER: US/10/632,793
; PRIOR FILING DATE: 2003-08-04
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: US/09/869,927
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: PCT/FR00/00144
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: FR 99/00888
; PRIOR FILING DATE: 1999-01-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 7582
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: misc feature
; LOCATION: (198)..(198)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (307)..(307)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (355)..(355)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1309)..(1309)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1331)..(1331)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2213)..(2213)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2398)..(2398)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3787)..(3787)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4115)..(4115)
; OTHER INFORMATION: n = a or g or c or t/u
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4261)..(4261)
; OTHER INFORMATION: n = a or g or c or t/u
US-10-632-793-30
```

```
Query Match 54.4%; Score 722.6; DB 16; Length 7582;
Best Local Similarity 88.5%; Pred. No. 1.le-213;
Matches 774; Conservative 19; Mismatches 70; Indels 12; Gaps 1;

QY 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCAAAAGAGGGGGAACCTGTTTATTTT 60
Db 6720 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCTGARAGGGGGGAACCTGTTTATTTT 6779
QY 61 AGGGGAAGAATGCTGTTAGTATGTTAATCAATCTGGAATCATTACTGAGAAAGTTAAAGA 120
Db 6780 AGGGGAAGAATGCTGTTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 6839
QY 121 AATTGAGATCGAATAATATGTAGCAGAGAGACCTTCAAAACACATGCACCTGGGGCCT 180
Db 6840 AATTGAGATCGAATAACACGTAACAGAGAGCTTCAAAACACATGCACCTGGGGCCT 6899
QY 181 CCTCAGCCATGATGCTGCTGGACTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATT 240
Db 6900 CCTCAGCCATGATGCTGCTGGACTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATT 6959
QY 241 TTACTCCTCTTTGGACCCCTGTATCTTCAACTTCCCTTTGTTAACTTTGTCTCTCCAGAA 300
Db 6960 GCTACTCCTCTTTGGACCCCTGTATCTTTRACCTCTCTTTGTTAACTTTGTCTCTCCAGAA 7019
QY 301 TGAAGCTGTAAAGCTCAAAATAGTTCTTCAAATGGAAACCCAGATGAGTCCATGACTAA 360
Db 7020 CGAAGCTGTAAACTA-----CAAATGGAGCCCAAGATGAGTCCCAAGACTAA 7067
QY 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGATGTTAAAGACATTGA 420
Db 7068 GATCTACCGCAGACCCCTGGACCGGCTGTAGCCCAAGCTGTTAGTCTGTTAATGACATCA 7127
```

QY 421 AGTCACCCCTCCGAGGAAATCTCAACTGCACAAACCCCTACTACTACTCTCAATTCAGTAGG 480
Db 7128 AGGACCCCTCTGAGGAAATCTCAGCTGCACAACTCTACTACGCCCAATTCAGGAGG 7187
QY 481 AAGCAGTTAGAGCAGTTGTGAGCAAACTCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
Db 7188 AAGCAGTTAGAGCGGTSCTGCGCAACCTCCCAACAGCAGTCTAGGTTTTCCTGTTGAGA 7247
QY 541 GGTGAGCTGAGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCCNAAAGCCT 600
Db 7248 TGGGGAGCTGAGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCCNAAAGCCT 7307
QY 601 ANCTGGGAGGTGACCGCACTCAATCTTAAACATGGGCTTCAACTAGCTCACACCCG 660
Db 7308 AGSTGGGAGGTGACCACTCCACCTTTAAACACGGGCTTCAACTAGCTCACACCTG 7367
QY 661 ACCAATCAGAGCTCACTAAATGCTAAATCAGGCAAAACAGGAGTAAAGCAATAGCC 720
Db 7368 ACCAATCAGAGCTCACTAAATGCTAAATCAGGCAAAACAGGAGTAAAGCAATAGCC 7427
QY 721 AATCATCTATTGCTGAGACAGGAGGAGGAGCAAGGATTTGGGATATATAAATCAGGCA 780
Db 7428 AATCATCTATTGCTGAGACAGGAGGAGGAGCAAGGATTTGGGATATATAAATCAGGCA 7487
QY 781 TTCAAGCCAGCAACAGACCCCTTTGGTCCCTCCCTCCCTATTTGATGGGAGCTCTGTTT 840
Db 7488 TTCAAGCCAGCAACAGACCCCTTTGGTCCCTCCCTCCCTATTTGATGGGAGCTCTGTTT 7547
QY 841 CACTCTATTCTACTCTATTAAATCATGCAACTGCA 875
Db 7548 CATGCTATTCTACTCTATTAAATCTTGCACCTGCR 7582

RESULT 13
US-10-632-793-25
; Sequence 25, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: MALLET, Francois
; APPLICANT: VOISSET, Cecile
; TITLE OF INVENTION: ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
; FILE REFERENCE: 110048
; CURRENT APPLICATION NUMBER: US/10/632,793
; PRIOR FILING DATE: 2003-08-04
; PRIOR FILING DATE: 2000-01-21
; PRIOR FILING DATE: 2000-01-21
; PRIOR FILING DATE: 1999-01-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 1136
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-632-793-25

Query Match 54.3%; Score 721.2; DB 16; Length 1136;
Best Local Similarity 90.0%; Pred. No. 1.1e-213;
Matches 787; Conservative 0; Mismatches 75; Indels 12; Gaps 1;
QY 1 TCAAAATCGAAGAGCTTTAGACTTGTACCGGCAAAAGAGGGGAAACCTGTTATTTT 60
Db 254 TCAAAATCGAAGAGCTTTAGACTTGTACCGGCAAAAGAGGGGAAACCTGTTATTTT 313
QY 61 AGGGAGAGATGCTGTAGTATGTTAATCAATCGGAATCATCTAGAGAAAGTTAAGA 120
Db 314 AGGGAGAGATGCTGTAGTATGTTAATCAATCGGAATCATCTAGAGAAAGTTAAGA 373
QY 121 AATTGAGATCGAATATATGTTAGAGCAGAGGACCTTCAAAACACTGCACCCCTGGGCGCT 180

Db 374 AATTCAGATCGAATACAACTAGACGACGAGAGCTTCGAAACCTTGGACCTTGGGCGCT 433
QY 181 CTTACGCCAATGGAATCCCTGGAGCTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATATT 240
Db 434 CTTACGCCAATGGAATCCCTGGAGCTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATATT 493
QY 241 TTTACTCTCTTTTGGACCCCTGTATCTTCAACTCTCTTCTTAACTTTGTCTCTTCCAGAA 300
Db 494 GCTACTCTCTTTTGGACCCCTGTATCTTAACTCTCTTCTTAACTTTGTCTCTTCCAGAA 553
QY 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAAACCCAGATGCAGTCCCATGACTAA 360
Db 554 CGAAGCTGTAAACTA-----CAATGGAGCCCAAGATGCAGTCCCAAGACTAA 601
QY 361 AATCTACCTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGTTTAAATGACATGA 420
Db 602 GATCTACCGCAGACCCCTGGACCGGCTGCTAGCCCAAGATGATGTTTAAATGACATCA 661
QY 421 AGTCACCCCTCCGAGGAAATCTCAACTGCACAAACCCCTACTACTACTCCAAATTCAGTAGG 480
Db 662 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACTCTACTACGCCCAATTCAGCAGG 721
QY 481 AAGCAGTTAGAGCAGTTGTGAGCAACCTCCCAACAGTACTTGGGTTTTCCTGTTGAGA 540
Db 722 AAGCAGTTAGAGCGGCTGCTGCGCAACCTCCCAACAGCAGCTTAGGTTTTCCTGTTGAGA 781
QY 541 GGTGAGCTGAGAGACAGGACTAGCTGGATTTTCTTAGGCTGACTAAGAAATCCCNAAAGCCT 600
Db 782 TGGGGAGCTGAGAGACAGGACTAGCTGGATTTTCTTAGGCTGACTAAGAAATCCCNAAAGCCT 841
QY 601 ANCTGGAAAGGTGACCGCATCCATCTTTAAACATGCGGCTTGAACCTTAGCTCACACCG 660
Db 842 AGCTGGAAAGGTGACCATCCACCTTTAAACACAGCGGCTTGAACCTTAGCTCACACCTG 901
QY 661 ACCAATCAGAGAGCTCACTAAATGCTATCAGGCAAAACAGGAGTAAAGCAATAGCC 720
Db 902 ACCAATCAGAGAGCTCACTAAATGCTATCAGGCAAAACAGGAGTAAAGCAATAGCC 961
QY 721 AATCATCTATTGCTGAGAGCAGCGGGAAGCAAGGATTTGGGATATATAAATCTCAGGCA 780
Db 962 AATCATCTATTGCTGAGAGCAGCGGGAAGCAAGGATTTGGGATATATAAATCTCAGGCA 1021
QY 781 TTCAAGCCAGCAACAGCAACCCCTTTGGTCCCTCCCTCCCTATTTGATGGGAGCTCTGTTT 840
Db 1022 TTCAAGCCAGCAACAGCAACCCCTTTGGTCCCTCCCTCCCTATTTGATGGGAGCTCTGTTT 1081
QY 841 CACTCTATTCTACTCTATTAAATCATGCAACTGC 874
Db 1082 CATGCTATTCTACTCTATTAAATCTTGCAGCTGC 1115

RESULT 14
US-10-632-793-26
; Sequence 26, Application US/10632793
; Publication No. US20040048298A1
; GENERAL INFORMATION:
; APPLICANT: PARANHOS-BACCALA, Glaucia
; APPLICANT: MALLET, Francois
; APPLICANT: VOISSET, Cecile
; TITLE OF INVENTION: ENDOGENEOUS NUCLEIC ACID FRAGMENT ASSOCIATED WITH AN
; FILE REFERENCE: 110048
; CURRENT APPLICATION NUMBER: US/10/632,793
; PRIOR FILING DATE: 2003-08-04
; PRIOR FILING DATE: 2001-10-22
; PRIOR FILING DATE: 2000-01-21
; PRIOR FILING DATE: 1999-01-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1

```
; SEQ ID NO 26
; LENGTH: 2782
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-632-793-26

Query Match      53.6%; Score 712.6; DB 16; Length 2782;
Best Local Similarity 89.4%; Pred. No. 8.7e-211;
Matches 782; Conservative 0; Mismatches 81; Indels 12; Gaps 1;

1  TCAAAATCGAAGAGCTTTAGACTTGTCTAACCAGCAAGAGGGGAACTGTTTATTTT 60
Db  1902  TCGAAATCGAAGAGCTTTAGACTTGTCTAACCAGCAAGAGGGGAACTGTTTATTTT 1961

61  AGGGGAAGATGCTGTAGTATGTTAATCAATCTGGAATCAATCTGAGAAAGTTAAAGA 120
Db  1962  AGGGGAAGATGCTGTATATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTGAGA 2021

121  AATTGAGATCGAATATAATGTAGAGCAGAGACCTTCAAAAACATGCAACCTGGGGCCT 180
Db  2022  AATTCCAGATCGAATAACAAGTATAGCAGAGAGCTTCAAAAACATGCAACCTGGGGCCT 2081

181  CCTCAGCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db  2082  CCTCAGCGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2141

241  TTACTCTCTTTGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
Db  2142  GCTACTCTCTTTGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2201

301  TGAAGCTGTAAAGCTACAAATAGTCTTCAATGGAAACCCAGATGCAATGCAATGCTAA 360
Db  2202  CGAAGCTGTGAAACTA-----CAAAATGGAGCCCAAGATGCAATGCAATGCTAA 2249

361  AATCTACCTGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Db  2250  GATCTACGGAAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2309

421  AGTCACCTCTCCGAGGAAATCTCAACTGCAACACCCCTACTACACTCACTCAATTCAGTAGG 480
Db  2310  AGGCACCTCTCTGAGGAAATCTCAGCTGCAACACCTCTACTACGCCCCAATTCAGCAGG 2369

481  AAGCAGTTAGAGCAGTGTGAGCAACCTCCCAACAGTACTTGGGTTTCTGTTGAGA 540
Db  2370  AAGCAGTTAGAGCAGTGTGAGCAACCTCCCAACAGTACTTGGGTTTCTGTTGAGA 2429

541  GGGTGAATGAGAGCAGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db  2430  TGGGGGACTGAGAGCAGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2489

601  ANCTGGGAAGGTGACCGCATCTTAAACATGCGGCTTGCACCTTACTGCTCAACCTCAGCCCG 660
Db  2490  AGGTGGGAAGGTGACCGCATCTTAAACATGCGGCTTGCACCTTACTGCTCAACCTCAGCCCG 2549

661  ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAACAGAGAGTTAAAGCAATAGCC 720
Db  2550  ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAACAGAGAGTTAAAGCAATAGCC 2609

721  AATCATCTATTGCTGAGAGCAGCGGAGAGCAAGGATGCGGATATAAATCAGGCA 780
Db  2610  AATCATCTATTGCTGAGAGCAGCGGAGAGCAAGGATGCGGATATAAATCAGGCA 2669

781  TTCAAGCCAGCAACAGCCCTTTGGGTCCTCCCTCCCAATGCTGAGGAGCTCTGTTT 840
Db  2670  TTGAGCCGCAACAGCCCTTTGGGTCCTCCCTCCCAATGCTGAGGAGCTCTGTTT 2729

841  CACTCTATTTCATCTTATTAATCATGCAACTGCA 875
Db  2730  CATGCTATTTCATCTTATTAATCATGCAACTGCA 2764

RESULT 15
US-10-133-036-1
```

```
; Sequence 1, Application US/10133036
; Publication No. US20040054133A1
; GENERAL INFORMATION:
; APPLICANT: Conrad, Bernard
; TITLE OF INVENTION: Multiple Sclerosis-Related Superantigen
; FILE REFERENCE: 23135-507
; CURRENT APPLICATION NUMBER: US/10/133,036
; PRIOR FILING DATE: 2002-04-26
; PRIOR APPLICATION NUMBER: PCT/EP00/10659
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2782
; TYPE: DNA
; ORGANISM: Human endogenous retrovirus
US-10-133-036-1

Query Match      53.5%; Score 711; DB 16; Length 2782;
Best Local Similarity 89.3%; Pred. No. 2.7e-210;
Matches 781; Conservative 0; Mismatches 82; Indels 12; Gaps 1;

1  TCAAAATCGAAGAGCTTTAGACTTGTCTAACCAGCAAGAGGGGAACTGTTTATTTT 60
Db  1902  TCGAAATCGAAGAGCTTTAGACTTGTCTAACCAGCAAGAGGGGAACTGTTTATTTT 1961

61  AGGGGAAGATGCTGTAGTATGTTAATCAATCTGGAATCAATCTGAGAAAGTTAAAGA 120
Db  1962  AGGGGAAGATGCTGTATATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTGAGA 2021

121  AATTGAGATCGAATATAATGTAGAGCAGAGACCTTCAAAAACATGCAACCTGGGGCCT 180
Db  2022  AATTCCAGATCGAATAACAAGTATAGCAGAGAGCTTCAAAAACATGCAACCTGGGGCCT 2081

181  CCTCAGCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db  2082  CCTCAGCGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2141

241  TTACTCTCTTTGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
Db  2142  GCTACTCTCTTTGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2201

301  TGAAGCTGTAAAGCTACAAATAGTCTTCAATGGAAACCCAGATGCAATGCAATGCTAA 360
Db  2202  CGAAGCTGTGAAACTA-----CAAAATGGAGCCCAAGATGCAATGCAATGCTAA 2249

361  AATCTACCTGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Db  2250  GATCTACGGAAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2309

421  AGTCACCTCTCCGAGGAAATCTCAACTGCAACACCCCTACTACACTCACTCAATTCAGTAGG 480
Db  2310  AGGCACCTCTCTGAGGAAATCTCAGCTGCAACACCTCTACTACGCCCCAATTCAGCAGG 2369

481  AAGCAGTTAGAGCAGTGTGAGCAACCTCCCAACAGTACTTGGGTTTCTGTTGAGA 540
Db  2370  AAGCAGTTAGAGCAGTGTGAGCAACCTCCCAACAGTACTTGGGTTTCTGTTGAGA 2429

541  GGGTGAATGAGAGCAGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db  2430  TGGGGGACTGAGAGCAGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2489

601  ANCTGGGAAGGTGACCGCATCTTAAACATGCGGCTTGCACCTTACTGCTCAACCTCAGCCCG 660
Db  2490  AGGTGGGAAGGTGACCGCATCTTAAACATGCGGCTTGCACCTTACTGCTCAACCTCAGCCCG 2549

661  ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAACAGAGAGTTAAAGCAATAGCC 720
Db  2550  ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAACAGAGAGTTAAAGCAATAGCC 2609

721  AATCATCTATTGCTGAGAGCAGCGGAGAGCAAGGATGCGGATATAAATCAGGCA 780
Db  2610  AATCATCTATTGCTGAGAGCAGCGGAGAGCAAGGATGCGGATATAAATCAGGCA 2669

781  TTCAAGCCAGCAACAGCCCTTTGGGTCCTCCCTCCCAATGCTGAGGAGCTCTGTTT 840
Db  2670  TTGAGCCGCAACAGCCCTTTGGGTCCTCCCTCCCAATGCTGAGGAGCTCTGTTT 2729

841  CACTCTATTTCATCTTATTAATCATGCAACTGCA 875
Db  2730  CATGCTATTTCATCTTATTAATCATGCAACTGCA 2764

RESULT 15
US-10-133-036-1
```

Db 2610 AATCATTTATTGCTTGAGACACAGCAGGAGGACAAATGATCGGATATAAAACCAAGTT 2669
QY 781 TTCAAGCCAGCAACAGCAACCCCTTTGGTCCCTCCCATTTGATGGAGCTCTGTTTT 840
Db 2670 TTGAGCCGCAACAGGCAACCCCTTTGGTCCCTCCCTTTGATGGAGCTCTGTTTT 2729
QY 841 CACTCTATTTCACCTATTAAATCATGCAACTGCA 875
Db 2730 CATGCTATTTCACCTATTAAATCTTGCAACTGCA 2764

RESULT 16

US-09-864-761-4444
; Sequence 4444, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aomic-X-1
; CURRENT APPLICATION NUMBER: US/09/864, 761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 4917
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4444
; LENGTH: 1894
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002346.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
; OTHER INFORMATION: EXPRESSED IN HELL100, SIGNAL = 6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9

; OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
US-09-864-761-4444
Query Match 53.0%; Score 704.8; DB 9; Length 1894;
Best Local Similarity 89.7%; Pred. No. 1.9e-208;
Matches 770; Conservative 0; Mismatches 79; Indels 9; Gaps 1;
QY 1 TCAAAATCGAAGAGCTTTAGACTTGCTAAACCGCCAAAGAGGGGGAACCTGTTTATTTTT 60
Db 961 TCAAAATCGAAGAGCTTTAGACTTGCTAAACCGGAAAGCGGGGAACCTTTTTATTTTT 1020
QY 61 AGGGGAAGAAATGCTGTGTAGTATGTTAAATCAATCTCGAATCATTTACTGAGAAAGTTAAAGA 120
Db 1021 AGAGGAAATGCTGTGTGTATGTTAAATCAATCCGGAATCATCCCGAAGAAAGTTAAAGA 1080
QY 121 AATTTGAGATCGAATAATAATGTAGACGAGAGACCTTCAAAAACATCTGCACCCCTGGGCCT 180
Db 1081 AATTCAGGTGGAATATAACGTAGACAAAGAGAGTGCAAAACACTGGACCCCTGGGCCT 1140
QY 181 CCTCAGCAATGGATGCGCTGGACTCTCCCTCTTTAGGACCTCTAGCAGCTATAATTT 240
Db 1141 CCTCAGCAATGGATGCGCTGGACTCTCCCTCTTTAGGACCTCTAGCAGCTATAATTT 1200
QY 241 TTTTACTCTCTTTGGACCTGTATCTTCAACTTCTTTAAAGTTTGTCTCTTCCAGAAT 300
Db 1201 GTTACTCTCTTTGGACCTGTATCTTAACTCTCTTTAAAGTTTGTCTTTCAGAAAT 1260
QY 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAACCCAGATGCAGTCCATGACTAA 360
Db 1261 CGAAGCAGTAAACTACAAATGTTCTTCAAAATGGAGCCCGACAGTGCAGTCCATGACTAA 1320
QY 361 AATCTACCGTGAACCCCTGGACCGGCTGTAGACTATGCTCTGATGTTTAATGACATGA 420
Db 1321 AATCTACCGTGAACCCCTGGACCGGCTGTAGACTATGCTCTGATGTTTAATGACATGA 1380
QY 421 AGTCACCCCTCCGAGGAAATCTCAACTGCACAAACCCCTACTACACTTCAAAATTCAGTAGG 480
Db 1381 AGGCACCCCTCCGAGGAAATCTCAACTGCACAAACCCCTACTACACTTCAAAATTCAGTAGG 1440
QY 481 AAGCAGTTAGACAGTGTGTGAGCCAACTCCCAACAGTACTTTGGGTTTCTCTGTTGAGA 540
Db 1441 AAGCAGTTAGAGTGTGTGAGCCAACTCCCAACAGTACTTTGGGTTTCTCTGTTGAGA 1500
QY 541 GGGTGGACTGAGACAGGACTAGCTGGATTTTCTTAGGCTGACTAAGAAATCCCNAGCCT 600
Db 1501 GGGGGGACTGAGACAGGAAATAAATAGATTTTCTAGACCAACTAAGAAATCCCTAAGACT 1560
QY 601 ANCTGGAGAGTGCACCGCATCCATCTTTAAACATGGGGTTCGAACCTTAGCTCACACCCG 660
Db 1561 AGCTGGAGAGTGCACCGCATCCATCTTTAAACATGGGGTTCGAACCTTAGCTCACACCCG 1620
QY 661 ACCAATC-----AGAGAGCTCACTAAATGCTTAATCAGGCCAAACAGGAGGTAAA 711
Db 1621 ACCAATCAGTACTAAAGAGAGCTCACTAAATGCTTAATAGGCCAAACAGGAGATAA 1680
QY 712 GCAATAGCCAATCATCTATTGCTGTAGAGCAGACGGGGAAGGACAAGGATTTGGGATATAA 771
Db 1681 GAAATAGCCAATCATCTATTGCTGTAGCAGACAGCAGGAGGGAACAATGATCGGATATAA 1740
QY 772 ACTCAGGCAATCAAGCCAGCAACAGCAACCCCTTTGGGTCCTCCCTCCCTATGTTATGGAG 831
Db 1741 ACCAGGCAATCGAGCCAGCTACAGCTACCTCTTTGGGTCCTCCCTCCCTTTGTTATGGAG 1800
QY 832 CTCTGTTTTCACTCTATT 849
Db 1801 CTCTGTTTCACTCTATT 1818

RESULT 17

US-10-087-192-730/c
; Sequence 730, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W. K.
; APPLICANT: Engelhardt, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 730
; LENGTH: 161334
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(161334)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-730

Query Match 48.4%; Score 643.4; DB 13; Length 161334;
Best Local Similarity 85.4%; Pred. No. 3e-188; Indels 46; Gaps 4;
Matches 781; Conservative 0; Mismatches 88;
QY 1 TCAAAATCGAAGAGCTTTAGACTGTCTAACCGCAAAAGAGGGGGAACCTGTTTATTTT 60
DB 100898 TCAAAATGAGAGCTTTAGACTGTCTAACCTCGAAGAGGGGGAAGCTGTTTATTTT 100839
QY 61 AGGGGAAGAAATGCTGTAG-----TATGTTAATCA 90
DB 100838 AGGGGAAGAAATGTTTATTTATGTTATTTTATGCGGAAGAAATGTTTATTTATGTTAATCA 100779
QY 91 ATCTGGAATCATCTACTGAGAACTTAAAGAAATTCAGATCGAATATAATGTAGACAGA 150
DB 100778 ATCTGGAATGTGACAGAGAAATGGAAGAAATTCAGATGTAATACAGGTAGAACAGA 100719
QY 151 GGACCTTC-AAAACATCGACACCTCGGGCCCTCTCAGCCAAATGGATGCCCTGGACTCTCC 209
DB 100718 GGAGCTTCAAAACACAGACCTCGGGCCCTCTCAGCCAAATGGATGCCCTGGACTCTCC 100659
QY 210 CCTCTTAGGACCTCTAGCAGCTATATATTTTATCTCTCTTTGACCCCTGTATCTCA 269
DB 100658 CCTCTTAGGATCTCTAGCAGCTCTAATATTGATCTCTCTTTGGACCCCTGTATCTTTA 100599
QY 270 ACTTCCTGTTAAGTTGTCTCTTCAGAAATGAAGCTGTAAGCTGTAAGCTGTAAGCTGTTCTTC 329
DB 100598 ACTTCCTGTTAAGTTGTCTCTTCAGAAATGAAGCTGTAAGCTGTAAGCTGTAAGCTGTTCTTC 100539
QY 330 AAATGGAACCCAGAGTGCAGTCCATGACTAAGAAATCTACCGTGGACCCCTGGACCGGCTG 389
DB 100538 AAATGGAACCCAGAGTGCAGTCCATGACTAAGAAATCTACCGTGGACCCCTGGACCGGCTA 100479
QY 390 CTAGACTATGCTCTGATGTTAATGACATGGAATGACATGACCCCTCCGAGGAAATCTCAACTG 449
DB 100478 CTAGCCCATGCTCCAAATGTTAATGATATCGAAGCCACCCCTCCGAGGAAATCTCAACTG 100419
QY 450 CACAAACCCCTACTACCTCCAAATTCAGTAGAAGCAGTGTAGACAGTGTGTCAGCCAACT 509
DB 100418 CACAAACCCCTACTATGCCCCAAATTCGCGAAGAGCAGTGTAGACTGTTGTCGTCAGCCAACT 100359
QY 510 CCCCAACAGTACTTGGGTTTCTGTTGAGAGGCTGAGTGTAGAGACAGGACTAGCTGGA 569
DB 100358 CCCCAACAGTACTTGGGTTTCTGTTGAGTGGGGGACTGAGACAGGATTAGCTGGA 100299
QY 570 TTTCCTAGGCTGACTAAGAAATCCNAAGCCTANTCTGGGAAGGTGACCGCATCCATCTTTA 629

DB 100298 TTTCTAGGCGGACTAAGAAATCCAAAGCTTAGCTGGGAAGGTGACCAATCCACCTTTA 100239
QY 630 AACATGGGCTTGGCAACTTAGCTTACACCCGACCAATC-----AGAGAGCTCACTA 680
DB 100238 AACATGGGCTTGGCAACTTAGCTTACACCCGACCAATCAGGTAGTAAGAGAGCTCACTA 100179
QY 681 AAATGCTAATCAGGCAAAACAGAGGTAAGCAATAGCAATCATCTATTGCTGAGAG 740
DB 100178 AAATGCTAATCAGGCAAAACAGAGGTAAGCAATAGCAATCATCTATCGCTGAGAG 100119
QY 741 CACAGCGGGAAGGACAGGATTTGGGATATAAATCTCAGGCAATTCAGCCGACCAAGCAAC 800
DB 100118 CACAGCGGGAAGGACAGGATTTGGGATATAAATCCAGGCAATTCAGCCGCAAGGCTAC 100059
QY 801 CCCCTTTGGGTCCCTCCCATTTGATGGGAGCTCTGTTTCACTCTATTCTCTATT 860
DB 100058 CTTCTTTGGGTCCCTCCCTTTGATGGGAGCTCTCTCT-----GTCCTTCACTCTATT 100005
QY 861 AATCATGCAACTGCA 875
DB 100004 AATATTGCAACTGCA 99990

RESULT 18

US-09-854-867-385
; Sequence 385, Application US/09854867
; Publication No. US20030224356A1
; GENERAL INFORMATION:
; APPLICANT: JOAN, KNOLL H
; APPLICANT: ROGAN, PETER K
; TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING
; FILE REFERENCE: 30307
; CURRENT APPLICATION NUMBER: US/09/854,867
; CURRENT FILING DATE: 2003-05-08
; NUMBER OF SEQ ID NOS: 613
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 385
; LENGTH: 780
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: (1)..(780)
; OTHER INFORMATION: It+17
US-09-854-867-385

Query Match 48.0%; Score 637.8; DB 10; Length 780;
Best Local Similarity 90.0%; Pred. No. 9.5e-188; Indels 19; Gaps 2;
Matches 711; Conservative 0; Mismatches 60;
QY 549 TGAGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCNAAGCTTANCTGGGA 608
DB 1 TGAGAGACAGGACTAGCTGGATTTCTTAGGCGGACTAAGAAATCCCTAAGCTAGCTGGGA 60
QY 609 AGGTGACCGCATCCTTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATC- 667
DB 61 AGGTGACCGCTTCCACTTTTAAACACGGGGCTTGCAACTTAGCTCACACCCGACCAATCA 120
QY 668 -----AGAGAGCTCCTTAAATGCTAATCAGGCAAAACAGGAGGTAAAGCAATAGC 719
DB 121 GGTAGTAAAGAGAGCTCCTTAAATGCTAATTAGGCAAAACAGGAGGTAAAGAAATAGC 180
QY 720 CAATCATCTATTGCTCTGAGAGCAAGCGGGAAGCAAGGATTGGGATATAAATCTCAGGC 779
DB 181 CAATCATCTATCGCTGAGAGCAAGCGGGAAGCAATGATCGGATATAAACCAGGC 240
QY 780 ATTCAAGCAGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGATGGAGCTCTGTT 839
DB 241 ATTCGAGCGGCAACCGCTACCTTTTGGGTCCCTCCCTTTGATGGAGCTCTGT- 298
QY 840 TCACTCTATTCTCTCTTAAATCATGCAACTGCACTCTCTGCTCCGTTGTTTTTATG 899

Db 299 -----TTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCGCTGTTGTTAAG 350
QY 900 GCTCAAGCTGAGCTTTTGGTGGCATCCACACATGCTGTTGGTGGCATCCGTCACAGCCGC 959
Db 351 GCTCGAGCTGAGCTTTTGGTGGCATCCACACATGCTGTTGGTGGCATCCGTCACAGCCGC 410
QY 960 TGCTGACTTCCATCCCTTTTGGATCCAGCAGAGTGTCCACATGCTGCTCTGATCCAGCGAGG 1019
Db 411 CGCTGACTTCCATCCCTTTTGGATCCAGCAGAGTGTCCACATGCTGCTCTGATCCAGCGAGG 470
QY 1020 TACCCATTTGCTTCCATCCGATCAGGCTAAAGGCTTGGCAATTTTCTGTCATGGCTAAAGTGC 1079
Db 471 CGCCCATTTGCTTCCATCCGATCAGGCTAAAGGCTTGGCAATTTTCTGTCATGGCTAAAGTGC 530
QY 1080 CTGGGTTTGTCTTAATAGAACTGAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1139
Db 531 CTGGGTTTGTCTTAATAGAACTGAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 590
QY 1140 CCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTTCCATTTCTTGGT 1199
Db 591 CCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTTCCATTTCTTGGT 650
QY 1200 ATCTGTGAGGCCAAGAACCCAGGCTCAGAGAANGTGGGCTTGGCCACCATTTGGGAAGTG 1259
Db 651 ATCTGTGAGGCCAAGAACCCAGGCTCAGAGAANGTGGGCTTGGCCACCATTTGGGAAGTG 710
QY 1260 GCCCACTGCAATTTTGTAGCGGCCACCAACCATTTTGGGAGCTGTGGGAGCAAGATCC 1319
Db 711 GCCCGCGCCATTTTGGAGCGGCCGCCACCATTTTGGGAGCTGTGGGAGCAAGATCC 770
QY 1320 CCAGTAACA 1329
Db 771 CCGGTAACA 780

RESULT 19

US-10-672-764A-34/c

; Sequence 34, Application US/10672764A

; Publication No. US20040156832A1

; GENERAL INFORMATION:

; APPLICANT: Jolly, Chris

; TITLE OF INVENTION: Immunoglobulin Compositions and Methods

; FILE REFERENCE: 1331.10010

; CURRENT APPLICATION NUMBER: US/10/672, 764A

; CURRENT FILING DATE: 2003-09-26

; NUMBER OF SEQ ID NOS: 68

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 34

; LENGTH: 157090

; TYPE: DNA

; ORGANISM: Human

US-10-672-764A-34

Query Match 45.3%; Score 601.4; DB 17; Length 157090;
Best Local Similarity 87.0%; Pred. No. 3.8e-175;
Matches 689; Conservative 0; Mismatches 84; Indels 19; Gaps 2;

QY 547 ACTGAGACAGCAGCTAGCTGGATTTCTAGGCTGACTAAGAATCCCNAAAGCCTACTGG 606
Db 81195 ATTGAGACAGCAGCTAGCTGGATTTCTAGGCTGACTAAGAATCCCNAAAGCCTACTGG 81136
QY 607 GAAGGTGACCGCATCCATCTTTAAACAATGGGGCTTGGCACTTAGCTCACACCCGACCAAT 666
Db 81135 GAAGGTGACCGCATCCATCTTTAAACAATGGGGCTTGGCACTTAGCTCACACCCGACCAAT 81076
QY 667 C-----AGAGCTCTAATAATGCTAATCAGGCAAAAACAGGAGTAAAGCATTA 717
Db 81075 CAGATAGTAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGTAAAGAAATA 81016
QY 718 GCCAATCATCTATTGCTGAGACACAGCGGAGGACAGGATTTGGGATATAAAGCTCAG 777
Db 81015 GCCAATCATCTATTGCTGAGACACAGCGGAGGACAGGATTTGGGATATAAAGCTCAG 80956

QY 778 GCATTTCAAGCCAGCAACAGCAACCCCTTTGGGTTCCCTCCCATTTGATATGGAGCTCTGT 837
Db 80955 GCATTTCAAGCCAGCAACAGCAACCCCTTTGGGTTCCCTCCCATTTGATATGGAGCTCTGT 80896
QY 838 TTTCACTCTATTATTTCACTCTATTAAATCATCAAACTGCACATCTTTCTGGTCCGTTTATTA 897
Db 80895 -----CTTCACTCTACTAAATCTTGCACATGCACTCTTCTGGTCCATGTTGTTA 80846
QY 898 TGCTCAAGCTGAGCTTTTGTTCGCATCCACACATGCTGTTTGGCCACCGTCACAGACCC 957
Db 80845 CGGCTCGAGCTTTCGCTTTCGCTCACCGTCCACCACTGCTGTTTGGCCCGCGTGGAGACCC 80786
QY 958 GCTGCTGACTTTCATCCCTTTTGGATCCAGCAGAGTGTCCCACTGCTGCTGATCCACGCA 1017
Db 80785 ACCGCTGGGCTTCCATCCCTCCGATCCATCAGGCTATCCGCTGCTGCTGATCCACGCA 80726
QY 1018 GGTACCCATTTGCCACTCCCGATCAGGCTAAAGGCTTGGCAATTTTCTGTCATGGCTAAGT 1077
Db 80725 GGGCCCATTTGCCACTCTCGATCGGCAAAATGCTGCCAATTTTCTGTCACGGCTAAGT 80666
QY 1078 GCTGGGTTTGTCTTAATAGAACTGAACATGCTGCTACTGGGTTCCATGTTCTCTTCAT 1137
Db 80665 GCCTGGGTTTGTCTTAATAGAACTGAACATGCTGCTACTGGGTTCCACGGTTCTCTCCGT 80606
QY 1138 GACCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTTCCATTTCTTG 1197
Db 80605 GACCCACGGCTTCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTTCCATTTCTTG 80546
QY 1198 GTATCTGTGAGGCCAAGAACCCAGGCTCAGAGAANGTGGGCTTGGCCACCATTTGGGAAG 1257
Db 80545 GAATCTGTGAGGCCAAGAACCCAGGCTCAGAGAANGTGGGCTTGGCCACCATTTGGGAAG 80486
QY 1258 TGGCCCACTGCTCAATTTTGTGAGCGGCCACCAACCATTTTGGGAGCTGTGGGAGCAAGAT 1317
Db 80485 TGGCTTGGGCTCAATTTTGGGAAGCGGCTGCAACCATTTTGGGAGCTGTGGGAGCAAGAT 80426
QY 1318 CCCCCAGTAACA 1329
Db 80425 CCCCCAATAACA 80414

RESULT 20

US-10-003-806-6/c

; Sequence 6, Application US/10003806

; Publication No. US20020119929A1

; GENERAL INFORMATION:

; APPLICANT: Bishop, Colin E.

; APPLICANT: Agoulinik, Alexander I.

; APPLICANT: Zhu, Qichao

; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY

; FILE REFERENCE: P02066US1/10024824

; CURRENT APPLICATION NUMBER: US/10/003, 806

; CURRENT FILING DATE: 2001-11-02

; PRIOR APPLICATION NUMBER: 60/245, 872

; PRIOR FILING DATE: 2000-11-03

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 6

; LENGTH: 180557

; TYPE: DNA

; ORGANISM: Human

US-10-003-806-6

Query Match 43.6%; Score 579.8; DB 13; Length 180557;
Best Local Similarity 86.0%; Pred. No. 2.2e-168;
Matches 683; Conservative 0; Mismatches 90; Indels 21; Gaps 3;

QY 547 ACTGAGACAGCAGCTAGCTGGATTTCTTAGGCTGACTAAGAATCCCNAAAGCCTACTGG 606
Db 58991 AGTGAGACAGCAGCTAGCTGGATTTCTTAGGCGGACTAAGAATCCCTAAGCCTAGCTGG 58932
QY 607 GAAGGTGACCGCATCCATCTTTAAACAATGGGGCTTGGCACTTAGCTCACACCCGACCAAT 666

Db 58931 GAAGGTGACCGCTTCCACCTTTAAACATGGGGCTTGCACCTTAGCTCAACACCGACCAAT 58872
QY 667 C-----AGAGAGCTCACTAAATGCTAATCAGGCAAAACAGGAGGTAAGCAATA 717
Db 58871 CAGATAGTAAGAGAGCTCACTAAATGCTAATTAGGCAAAACAGGAGGTAAGCAATA 58812
QY 718 GCCAATCATCTATTTGCTGAGAGCAGCGGGAGGAGCAAGGATTTGGGATATAAATCTAG 777
Db 58811 GCCAATCATCTATTTGCTGAGAGCAGCGGGAGGAGCAAGGATTTGGGATATAAATCTAG 58752
QY 778 GCATTCAAGCCAGCAACACCCCTTTGGGTCCCTCCCAATTTGATGGGAGCTCTGT 837
Db 58751 GCATTGAGCCAGCAACACCGCTTTGGGTCCCTCCCAATTTGATGGGAGCTCTGT 58692
QY 838 TTTCACCTCTATTTCACTCTAATAATCATGCAACTGCA--CTCTTCTGGTCCGTGTTTTT 895
Db 58691 -----CTTCACCTCTAATAATCATGCAACTGCA--CTCTTCTGGTCCGTGTTTTT 58642
QY 896 TATGGCTCAAGCTGAGCTTTTGTTCGCCATCCAGGCTTAAGGCTTGCCATTGTTCTGATGGCTAA 1075
Db 58641 CATGGTTCGAGCTGAGCTTTCTCTCGCGCTCCACCACTGCTGTTTGGCGCTGTGCGAGAC 58582
QY 956 CGCTGCTGACTTCCATCCGTCAGATCCGGAAGGTGCTGCTGTGCTCTGATCCAGC 1015
Db 58581 CTGCTGCTGACTTCCATCCGTCAGATCCGGAAGGTGCTGCTGTGCTCTGATCCAGC 58522
QY 1016 GAGGTACCAATGCCCACTCCGATCAGGCTTAAGGCTTGCCATTGTTCTGATGGCTAA 1075
Db 58521 GAGGTACCAATGCCCACTCCGATCAGGCTTAAGGCTTGCCATTGTTCTGATGGCTAA 58462
QY 1076 GTGCTGGGTTGCTCTAATAGAGCTATAACACTGCGGCTTCCACCTGCTGCTCTCTTCC 1135
Db 58461 GTGCTGGGTTGCTCTAATAGAGCTATAACACTGCGGCTTCCACCTGCTGCTCTCTTCC 58402
QY 1136 ATGACCCAGGCTTCTAATAGAGCTATAACACTGCGGCTTCCACCTGCTGCTCTCTTCC 1195
Db 58401 ATGACCCAGGCTTCTAATAGAGCTATAACACTGCGGCTTCCACCTGCTGCTCTCTTCC 58342
QY 1196 TGGTATCTGTAGGCGCAAGAACCCAGGTCAGAGAANGTGAGGCTTGGCACCATTGGGA 1255
Db 58341 TGGTATCTGTAGGCGCAAGAACCCAGGTCAGAGAANGTGAGGCTTGGCACCATTGGGA 58282
QY 1256 AGTGGCCCACTGCGCATTTTGGTAGCGGCCCAACCACTCTTGGGAGCTGTGGGAGCAAGG 1315
Db 58281 AGTGGCCCACTGCGCATTTTGGTAGCGGCCCAACCACTCTTGGGAGCTGTGGGAGCAAGG 58222

RESULT 21

US-10-003-806-9/c
; Sequence 9, Application US/10003806
; Publication No. US20020119929A1

GENERAL INFORMATION:

APPLICANT: Bishop, Colin E.

APPLICANT: Agulnik, Alexander I.

APPLICANT: Zhu, Qichao

TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY

FILE REFERENCE: P02066US1/10024824

CURRENT APPLICATION NUMBER: US/10/003, 806

CURRENT FILING DATE: 2001-11-02

PRIOR APPLICATION NUMBER: 60/245, 872

PRIOR FILING DATE: 2000-11-03

NUMBER OF SEQ ID NOS: 14

SOFTWARE: Patent.in version 3.1

SEQ ID NO 9

LENGTH: 180557

TYPE: DNA

ORGANISM: Human

US-10-003-806-9

Query Match 43.6%; Score 579.8; DB 13; Length 180557;
Best Local Similarity 86.0%; Pred. No. 2.2e-166;
Matches 683; Conservative 0; Mismatches 90; Indels 21; Gaps 3;

QY 547 ACTGAGAGACAGAGCTAGCTGGATTTCCTAGGCTGACTAAGAAATCCNNAAGCCTACTCTGG 606
Db 58991 AGTGAGAGACAGAGCTAGCTGGATTTCCTAGGCTGACTAAGAAATCCNNAAGCCTACTCTGG 58932
QY 607 GAAGGTGACCGCATCCATCTTTAAACATGGGCTTTGCAACTTAGCTTCAACCCGACCAAT 666
Db 58931 GAAGGTGACCGCTTCCACCTTTAAACATGGGCTTTGCAACTTAGCTTCAACCCGACCAAT 58872
QY 667 C-----AGAGAGCTCACTAAATGCTAATCAGGCAAAACAGGAGGTAAGCAATA 717
Db 58871 CAGATAGTAAGAGAGCTCACTAAATGCTAATCAGGCAAAACAGGAGGTAAGCAATA 58812
QY 718 GCCAATCATCTATTTGCTGAGAGCAGCGGGAGGAGCAAGGATTTGGGATATAAATCTAG 777
Db 58811 GCCAATCATCTATTTGCTGAGAGCAGCGGGAGGAGCAAGGATTTGGGATATAAATCTAG 58752
QY 778 GCATTCAAGCCAGCAACACCCCTTTGGGTCCCTCCCAATTTGATGGGAGCTCTGT 837
Db 58751 GCATTGAGCCAGCAACACCGCTTTGGGTCCCTCCCAATTTGATGGGAGCTCTGT 58692
QY 838 TTTCACCTCTATTTCACTCTAATAATCATGCAACTGCA--CTCTTCTGGTCCGTGTTTTT 895
Db 58691 -----CTTCACCTCTAATAATCATGCAACTGCA--CTCTTCTGGTCCGTGTTTTT 58642
QY 896 TATGGCTCAAGCTGAGCTTTTGTTCGCCATCCAGGCTTAAGGCTTGCCATTGTTCTGATGGCTAA 955
Db 58641 CATGGTTCGAGCTGAGCTTTCTCTCGCGCTCCACCACTGCTGTTTGGCGCTGTGCGAGAC 58582
QY 956 CGCTGCTGACTTCCATCCGTCAGATCCGGAAGGTGCTGCTGTGCTCTGATCCAGC 1015
Db 58581 CTGCTGCTGACTTCCATCCGTCAGATCCGGAAGGTGCTGCTGTGCTCTGATCCAGC 58522
QY 1016 GAGGTACCAATGCCCACTCCGATCAGGCTTAAGGCTTGCCATTGTTCTGATGGCTAA 1075
Db 58521 GAGGTACCAATGCCCACTCCGATCAGGCTTAAGGCTTGCCATTGTTCTGATGGCTAA 58462
QY 1076 GTGCTGGGTTGCTCTAATAGAGCTATAACACTGCGGCTTCCACCTGCTGCTCTCTTCC 1135
Db 58461 GTGCTGGGTTGCTCTAATAGAGCTATAACACTGCGGCTTCCACCTGCTGCTCTCTTCC 58402
QY 1136 ATGACCCAGGCTTCTAATAGAGCTATAACACTGCGGCTTCCACCTGCTGCTCTCTTCC 1195
Db 58401 ATGACCCAGGCTTCTAATAGAGCTATAACACTGCGGCTTCCACCTGCTGCTCTCTTCC 58342
QY 1196 TGGTATCTGTAGGCGCAAGAACCCAGGTCAGAGAANGTGAGGCTTGGCACCATTGGGA 1255
Db 58341 TGGTATCTGTAGGCGCAAGAACCCAGGTCAGAGAANGTGAGGCTTGGCACCATTGGGA 58282
QY 1256 AGTGGCCCACTGCGCATTTTGGTAGCGGCCCAACCACTCTTGGGAGCTGTGGGAGCAAGG 1315
Db 58281 AGTGGCCCACTGCGCATTTTGGTAGCGGCCCAACCACTCTTGGGAGCTGTGGGAGCAAGG 58222
QY 1316 ATCCCCCAGTAACA 1329
Db 58221 ACCACCTGGTAACA 58208

RESULT 22

US-10-017-117-1

; Sequence 1, Application US/10017117

; Publication No. US20030124535A1

GENERAL INFORMATION:

APPLICANT: McCarthy, Jeanette

TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE

FILE REFERENCE: MMI-008

CURRENT APPLICATION NUMBER: US/10/017,117

CURRENT FILING DATE: 2001-12-14

PRIOR APPLICATION NUMBER: No. US20030124535A1 yet assigned

PRIOR FILING DATE: 2001-12-05

; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 161671
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(161671)
; OTHER INFORMATION: n = a,t,c or g
US-10-017-117-1

Query Match 42.9%; Score 570.2; DB 15; Length 161671;
Best Local Similarity 88.9%; Pred. No. 2.1e-165; Indels 14; Gaps 3;
Matches 655; Conservative 0; Mismatches 71;
QY 547 ACTGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGATCCNAAAGCCCTACTGG 606
DB 74117 ACTGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGATCCCTAAGCTAGCTGG 74176
QY 607 GAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTTCACACCCGACCAAT 666
DB 74177 GAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTTCACACCCGACCAAT 74236
QY 667 CAGAGAGCTCACTAAATGCTAATCAGGCAAAACAGGAGGTAAAGCAATAGCCATCAT 726
DB 74237 CAGAGAGCTCACTAAATGCTAATCAGGCAAAACAGGAGGTAAAGCAATAGCCATCAT 74296
QY 727 CTATTCCTTGAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 786
DB 74297 CTATTCCTTGAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 74356
QY 787 CCAGCAACAGCAACCCCTTTGGGTCCTCCATGTTGATGGAGGCTCTGTTTCACTCT 846
DB 74357 CCGCAATGGCAACCCCTTTGGGTCCTCCATGTTGATGGAGGCTCTGTTTCACTCT 74407
QY 847 ATTCACTCTAATAATCATGCACTGCACTCTTCTGGTCCGCTGTTTATAGGCTCAAG 906
DB 74408 -TTTCACTCTAATAATCATGCACTGCACTCTTCTGGTCCGCTGTTTATAGGCTCAAG 74466
QY 907 CTGAGCTTTTGGCCATCCACCTGCTG-TTTCGCCCGTCACAGACCCGCTGCTGA 965
DB 74467 CTGAGCTTTTGGCCATCCACCTGCTGTTTGGCCCGTGGCAGACTCGCCACTGA 74526
QY 966 CTTTCATCCCTTTGGATCCAGCAGAGTGTCATGTCCTGATCCAGCAGAGGTACCCA 1025
DB 74527 CTTTCATCTTCGCGATCCGCGAGGCTGTCCTGCTGCTGATAGCAAGGCAACCA 74586
QY 1026 TTGCACTCCGATCAGGCTAA---AGGCTTGCCATGTTCTCTGATGGCTAAGTGCCTG 1082
DB 74587 CTGCCATCCGATCCGCTAAATTAAGGCTCGCATTTGTTCTGCTATGGCTAAGTGCCTG 74646
QY 1083 GGTTCCTCTAATAGAACTGACACTGTCACCTGGGTTCCATGGTCTCTTCCATGACCC 1142
DB 74647 GGTTCCTCTAATAGAACTGACACTGTCACCTGGGTTCCATGGTCTCTTCCGTCACCC 74706
QY 1143 ACGGCTTCTAATAGAGCTATACTCAGCGATGGCCCAAGATTCCATTCTTGGTATC 1202
DB 74707 ACGGCTTCTAATAGAGCTATACTCAGCGATGGCCCAAGATTCCATTCTTGGTATC 74766
QY 1203 TGTGAGCCCAAGAACCCAGGTCAGAGAANGTAGGCTTGCCACCAATTTGGGAAGTGGCC 1262
DB 74767 CGTGAGCCCAAGAACCCAGGTCAGAGAACAGGAGGCTTGTCACCATCTTGAAGCAGTC 74826
QY 1263 CACTGCCATTTGGTAGCG 1282
DB 74827 CACCGCCATCTTGAAGCTG 74846

RESULT 23

US-10-240-425-1102
; Sequence 1102, Application US/10240425
; Publication No. US20040033502A1

; GENERAL INFORMATION:
; APPLICANT: Williams, Amanda
; APPLICANT: Boland, Joseph F.
; APPLICANT: Lord, Reginald V.
; APPLICANT: Alvarez, Chris
; APPLICANT: Wetzel, Jon C.
; APPLICANT: Scherf, Uwe
; APPLICANT: Vockley, Joseph G.
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
; FILE REFERENCE: 44921-5026
; CURRENT APPLICATION NUMBER: US/10/240,425
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: PCT/US01/09847
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/193,446
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 1588
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1102
; LENGTH: 134292
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20040033502A1 AL031983
US-10-240-425-1102

Query Match 41.0%; Score 545.4; DB 16; Length 134292;
Best Local Similarity 85.9%; Pred. No. 1e-157;
Matches 680; Conservative 0; Mismatches 89; Indels 23; Gaps 6;
QY 547 ACTGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGATCCNAAAGCCCTACTGG 606
DB 54610 AATGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGATCCCTAAGCCCTAGCTGG 54669
QY 607 GAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTTCACACCCGACCAAT 666
DB 54670 GAAGGTGACTGATCCACCTTTAAACACGGGGCTTGCAACTTAGCTTCACACCCGACCAAT 54729
QY 667 C-----AGAGAGCTCACTAAATGCTAAATCAGGCAAAAAACAGGAGGTAAAGCAATA 717
DB 54730 CAGGTAGTAAAGAGAGCTCACTAAATGCTAAATAGGCAAAAAACAGGAGGTAAAGATATA 54789
QY 718 GCCATCATCTATTGCTGAGACAGCGGGAGGAGCAAGGATTTGGGATATAAATCTAG 777
DB 54790 GCCAATCATCTATTGCTGAGATCAAGCGGGAGGCAATGATCGGGATATAAACAAG 54849
QY 778 GCATTCAAGCAGCAACACCCCTTTGGGTCCTCCCATTTGATGGAGCTCTGT 837
DB 54850 GCATTGAGCAGCAAC-GCTACCTCTTTGGGTCCTCCCTTTGATGGAGCTCTGT 54908
QY 838 TTTCACTCTATTTCACCTCTATTAAATCATGCAACTGCACTCTTCTGCTCGTGTCTTTTA 897
DB 54909 -----CTTCACTCTATAAATCTTGAACCTGCACTCTTCTGCTCTATGTTCTTA 54958
QY 898 TGGCTCAAGCTGAGCTTTTGTTCGCCATCCACACCTGCTGTTTGGCCACCGTCAAGACCC 957
DB 54959 CGGCTGAGGCTGAGCTTTGCTTGGCATCCACCTGCGCTTTTGGCCACCGTCAAGACCC 55018
QY 958 GCTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGCTGCTCTGATCCAGCGA 1017
DB 55019 GCGGCTGACTTCCATCCCTTC-GGATCTGGAGGGTGTCCGCTGCTGCTCTGTTGAAACAGTGA 55077
QY 1018 GGTACCCATTGGCAGCTCCCGATCAGGCTAAAGGCTTTGCCATTTGTTCTGATGGCTAAGT 1077
DB 55078 GCGGCCATTGGCGCTCCTGATTGGGCTAAAGGCTAACGCGTACCATTGTT-CTGCAAGGCTAAGT 55136
QY 1078 GCTGGGTTTGTCTTAATAGAACTGAACACTGTGTCTCTGGGTTCCATGTTCTCTTCAT 1137
DB 55137 GCGGAGTTCTTCTTAATGAGCTGAACACTAGTCACTGGG-TCCACAGTCTCTCTTCCT 55195
QY 1138 GACCCACGCTCTTANTAGGCTATAAAGCTCAAGGCTGCGCCAGATTTCCATTCCTTG 1197
DB 55196 GACCCATGGCTCTTAATAGAGCTGTAAACCAACCCACACAGTCCAGATTTCCATTCCTTG 55255


```
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/979,847
/ FILING DATE: 26-NOV-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BERRIDGE, WILLIAM P.
/ REGISTRATION NUMBER: 30,024
/ REFERENCE/DOCKET NUMBER: WPB 39046A
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 703-836-6400
/ TELEFAX: 703-836-2787
/ INFORMATION FOR SEQ ID NO: 102:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 635 base pairs
/ TYPE: nucleic acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-10-114-104-102

Query Match      41.0%; Score 545; DB 15; Length 635;
Best Local Similarity 92.4%; Pred. No. 7.6e-159;
Matches 572; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 257 CCCTGATCTTCAACTCTCTGTTAAGTTTGTCTTCCAGAAATTGAAGCTGTAAGCTA 316
Db 1 CCCTGATCTTAACTCTCTGTTAAGTTTGTCTTCCAGAAATCAAAACTGTAAGCTA 60

QY 317 CAAATGATTTTCAATGGAACCCAGATGCAGTCATCACTAAATCTACCGTGGACCC 376
Db 61 CAAATGTTCTTCAATGAGCACCAGATGGAGTCATCACTAAATCACTACCGTGGACCC 120

QY 377 CTGGACCGGCTGTAGACTATGCTCTGATGTTAATGCAATTCAGTCACTACCGTCCCGAG 436
Db 121 CTGGACCGGCTGTAGACCATGCTCCGATGTTAATGCAATTCAGGACCATCTCCCGAG 180

QY 437 GAAATCTCAACTGCAACCCCTAATCACTCAATTCAGTGAAGAGAGTGTAGAGCACT 496
Db 181 GAAATCTCAACTGCAACCCCTAATCTATGCCCAATTCAGCGGAAGCAGTGTAGAGCGGT 240

QY 497 TGTACGCAACCTCCCAACAGTACTTGGTGTCTGTTGAGAGGGTGGACTGAGAGAC 556
Db 241 CATCAGCCCAACTCCCAACAGCAGTCTGGTGTCTGTTGAGAGGGGAGCTGAGAGAC 300

QY 557 AGGACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCAAGCTTANCTGGGAAGGTGACC 616
Db 301 AGGACTAGCTGATTTCTTAGGCCCAACGAGAAATCCCTAAGCTAGCTGGGAGGTGACT 360

QY 617 GCATCCATCTTTAAACATGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGTTC 676
Db 361 GCATCCACCTCTAAACATGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGTTC 420

QY 677 ACTAAATGCTTAATCAGGCAAAACAGGAGGTAAAGCAATAGCCATCATCTATTGCCCTG 736
Db 421 ACTAAATGCTTAATAGGCAAAATAGGAGGTAAAGAAATAGCCAATCATCTATTGCCCTG 480

QY 737 AGAGCACAGCGGAAGGACAAGGATGGGATATAAATCAGGCATTCAAGCCAGCAACAG 796
Db 481 AGAGCACAGCGGAGGACAAGGATCGGGATATAAACCAGGCATTTCGAGCGGCAACGG 540

QY 797 CAACCCCTTTGGGTCCCTCCCATTTGATGGAGCTCTGTTTCACTCTATTTCACCTCT 856
Db 541 CAACCCCTTTGGGTCCCTCCCTTTGATGGCGCTCTGTTTCACTCTATTTCACCTCT 600

QY 857 ATTAATCATGCACTGCA 875
Db 601 ATTAATCTTGCAACTGAA 619

RESULT 26
US-10-220-120-15
; Sequence 15, Application US/10220120
```

```
/ Publication No. US20040048253A1
/ GENERAL INFORMATION:
/ APPLICANT: INCYTE GENOMICS, INC.
/ APPLICANT: PANZER, Scott R.
/ APPLICANT: SPIRO, Peter A.
/ APPLICANT: BANVILLE, Steven C.
/ APPLICANT: SHAH, Purvi
/ APPLICANT: CHALUP, Michael S.
/ APPLICANT: CHANG, Simon C.
/ APPLICANT: CHEN, Alice
/ APPLICANT: D'SA, Steven A.
/ APPLICANT: AMSHEY, Stefan
/ APPLICANT: DAHL, Christopher R.
/ APPLICANT: DANIELS, Susan E.
/ APPLICANT: DUFOUR, Gerard E.
/ APPLICANT: FLORES, Vincent
/ APPLICANT: FONG, Willy T.
/ APPLICANT: GREENAWALT, Lila B.
/ APPLICANT: HILLMAN, Jennifer L.
/ APPLICANT: JONES, Anissa L.
/ APPLICANT: LIU, Tommy F.
/ APPLICANT: ROSEBERRY, Ann M.
/ APPLICANT: ROSEN, Bruce H.
/ APPLICANT: RUSSO, Frank D.
/ APPLICANT: STOCKREHER, Theresa K.
/ APPLICANT: DAFFO, Abel
/ APPLICANT: WRIGHT, Rachel J.
/ APPLICANT: YAP, Pierre E.
/ APPLICANT: YU, Jimmy Y.
/ APPLICANT: BRADLEY, Diana L.
/ APPLICANT: BRATCHEY, Shawn R.
/ APPLICANT: CHEN, Wensheng
/ APPLICANT: COHEN, Howard J.
/ APPLICANT: HODGSON, David M.
/ APPLICANT: LINCOLN, Stephen E.
/ APPLICANT: JACKSON, Stuart
/ TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: PT-1113 PCT
/ CURRENT APPLICATION NUMBER: US/10/220,120
/ CURRENT FILING DATE: 2002-08-26
/ PRIOR APPLICATION NUMBER: 60/184,777; 60/184,797; 60/184,698; 60/184,770; 60/184,774;
/ 60/184,693; 60/184,771; 60/184,813; 60/184,773; 60/184,776;
/ 60/184,769; 60/184,768; 60/184,837; 60/184,697; 60/184,841;
/ 60/184,772; 60/185,213; 60/185,216; 60/204,863; 60/205,221;
/ 60/204,815; 60/203,785; 60/204,821; 60/204,908; 60/204,226;
/ 60/204,525; 60/205,285; 60/205,232; 60/205,323; 60/205,287;
/ 60/205,324; 60/205,286
/ PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
/ 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
/ 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
/ 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
/ 2000-05-17; 2000-05-12; 2000-05-16; 2000-05-16; 2000-05-15;
/ 2000-05-16; 2000-05-17; 2000-05-17; 2000-05-17; 2000-05-17;
/ 2000-05-17; 2000-05-17
/ NUMBER OF SEQ ID NOS: 422
/ SOFTWARE: PERL Program
/ SEQ ID NO 15
/ LENGTH: 849
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No. US20040048253A1 LG:429446.1:2000FEB18
US-10-220-120-15

Query Match      40.3%; Score 536.2; DB 16; Length 849;
Best Local Similarity 81.4%; Pred. No. 5e-156;
Matches 650; Conservative 0; Mismatches 135; Indels 14; Gaps 2;

QY 208 CCCCTTCTTAGGACCTCTAGGAGCTATATATATTTTACTCTCTTTGGACCCCTGATCTT 267
Db 46 CACCATCTTGGGAGCTCTAAGAACCAAGACCCGCCGTAACATTTGGTGGCAACTGTACG 105
```

268 CAACCTTCCTTGTAAAGTTGCTCTCCAGAAATGAAGCTGTAAAGCTACAAATAGTTCT 327
106 GGAATTCCTCAAGTGTGTTGCTCTCCAGAAATCAAAAGCTGTAAAGCTACAAATAGTTCT 165
328 TCAATGGAAACCCAGATGAGTCCATAGTAAATATCCAGTGGGACCCCTGGACCGGCC 387
166 TCAATGGAGCCCGAGATGAGTCCATAGTAAATATCCAGTGGGACCCCTGGACCGGCC 225
388 TGCTAGACTATGCTCTGATGTTAATACATTTGAAGTCAACCTCCCGAGGAAATCTCAAC 447
226 TGCTAGCCCATGCTCCGATGTTGATGACATCGAAGGCGGCCCTCCCAAGAAATCTCAAC 285
448 TGCACAAACCCCTACTACACTCCAAATTCAGTAGGAAGCAGTTAGAGCAGTTGTCCAGCAAC 507
286 TGCTAGCCCTACTACAGCCCAATTCGCGAGGAGCAGTTAGAGCAGTTGTGGCCAAAC 345
508 CTCCCAACAGTACTCTGGGTTTCTGTTGAGAGGTGAGTCCAGAGCAGGACTAGTGG 567
346 CTCCCAACAGCAGTCTGAGTTTCTGTTGAGAGGAGAACTGAGAGCAGGACTAGTGG 405
568 GATTTCTAGGCTGACTAGATCCCAAGCCTTANCTGGGAGGTGACCGCATCCATCTT 627
406 GATTTCTAGGCTGACTAGATCCCAAGCCTTANCTGGGAGGTGACCGCATCCATCTT 465
628 TAAACATGGGCTTGCAACTTAGCTCACACCCCGACCAATCA-----GAGAGCTCAC 678
466 TAAACATGGGCTTGCAACTTAGCTCACACCCCGACCAATCA-----GAGAGCTCAC 525
679 TAAATGCTAATCAGCAAAACAGAGGTTAAGCAATAGCCAAATCATCTATTGCTCTGAG 738
526 TAAATGCTAATTAGGCAAAACAGAGGTTAAGCAATAGCCAAATCATCTATTGCTCTGAG 585
739 AGCAGAGCGGAGGACAGAGTTGGATATAAATCAAGCTTCAAGCCAGCAACAGCA 798
586 AGCAGAGCGGAGGACAGAGTTGGATATAAATCAAGCTTCAAGCCAGCAACAGCA 645
799 ACCCTCTTTGGGTCCTCCATTTGATGGAGCTCTGTTTTCATCTCTATTTCACTCTAT 858
646 ACCCTCTTTGGGTCCTCCGTTGATGGAGCTCTGTTTTCATCTCTATTTAAAGTCT 705
859 TAAATCATGCACTGCACTCTCTGCTCGGTGTTTATAGGCTCAAGCTGAGCTTTGT 918
706 TGAAC-----TGACACTCTCTGATCCATGTTGTTACAGCTTGAGCTGAGCTTTGC 760
919 TGCCATCCACACTGCTGTTTCCACCGCTCACAGACCCGCTGCTGACTTCCATCCCTTT 978
761 TCACCATCCAGCAATGCTCTTTGCCCACTCACAGACCCCGCACTGACTTCCATCCCTCT 820
979 GGATCCAGCAGAGTGTCCA 997
821 GGATCCAGCAGAGTGTCTCA 839

RESULT 27

US-10-719-993-6792/c
; Sequence 2833, Application US/10719993
; Publication No. US20040265849A1

GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6792

; LENGTH: 366710

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(366710)

; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-7
US-10-719-993-6792

Query Match 39.9%; Score 530.6; DB 18; Length 366710;

Best Local Similarity 91.0%; Pred. No. 7.3e-153; Indels 0; Gaps 0;
Matches 563; Conservative 0; Mismatches 56;

QY 257 CCTGTATCTTCAACCTTCCTTGTAAAGTTGCTCTCCAGAAATTTGAAGCTGTAAAGCTA 316
DB 43899 CCTGTATCTTAAACACCTTGTAAAGTTGCTCTCCAGAAATTTGAAGCTGTAAAGCTA 43840
QY 317 CAAATAGTCTTCAATGGAACCCAGATGAGTCCATGACATAAATCTTACCGTGGACCC 376
DB 43839 CAAATGTTCTTCAATGGAACCCAGATGAGTCCATGACATAAATCTTACCGTGGACCC 43780
QY 377 CTGGACCGGCTGCTAGATGATCTGCTGATGTTAATGACATTCAGCTCACCCCTCCCGAG 436
DB 43779 CTAGACTGCTGCTAGCCCATGCTCTGGTGTAAATGACATCGAAGTCACTCTCTCTGAG 43720
QY 437 GAAATCTCAACTGCAACACCCCTTACTACACTCCAAATTCAGTAGGAAGCAGTTAGAGCAGT 496
DB 43719 GAAATCTCAACTGCAACACCCCTTACTATGCCAGTTTCAGCAGGAAGCAGTTAGAGTGGT 43660
QY 497 TGTACGCCAACCCTCCCAACAGTACTTGGGTTTTCTGTTGAGAGGTGAGTGGAGAC 556
DB 43659 CATAGCCAAACCTTACCAACAGCAGTCTGGGTTTTCTGTTGAGAGGAGGAGTGGAGAC 43600
QY 557 AGGACTAGCTGGATTTCTTAGCTGACTAAGATCCCAAGCTTACTGGAAGGTGACC 616
DB 43599 AGGACTAGCTGGATTTCTTAGCTGACTAAGATCCCAAGCTTACTGGAAGGTGACC 43540
QY 617 GCATCCATCTTTAAACATGGGCTTGCAACTTAGCTCACACCCGCAACATTCAGAGAGCTC 676
DB 43539 GCATCCATCTTTAAACATGGGCTTGCAACTTAGCTCACACCCGCAACATTCAGAGAGCTC 43480
QY 677 ACTAAATGCTAATCAGGCAAAACAGAGGTAAAGCAATAGCCATCATCTATTGCTG 736
DB 43479 ACTAAATGCTAATCAGGCAAAACAGAGGTAAAGCAATAGCCATCATCTATTGCTG 43420
QY 737 AGGACACGCGGAGGAGCAAGATTTGGGATATAAATCAGGCTTCAAGCCAGCAACAG 796
DB 43419 AGGACACGCGGAGGAGCAAGATTTGGGATATAAATCAGGCTTCAAGCCAGCAACAG 43360
QY 797 CAACCCCTTTGGGTCCTCCATTTGATGGAGCTCTGTTTTCATCTATTTCATCTCT 856
DB 43359 CAACCCCTTTGGGTCCTCCATTTGATGGAGCTCTGTTTTCATCTATTTCATCTCT 43300
QY 857 ATTAATCATGCAACTGCA 875
DB 43299 ATTAATCATGCAACTGCA 43281

RESULT 28

US-10-264-237-2833/c
; Sequence 2833, Application US/10264237
; Publication No. US20040009491A1

GENERAL INFORMATION:

; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P4131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2833
; LENGTH: 17758
; TYPE: DNA
; ORGANISM: Homo sapiens

US-10-264-237-2833

Query Match	38.5%	Score 512	DB 16	Length 17758
Best Local Similarity	84.7%	Pred. No. 9.1e-148		
Matches 614	Conservative 0	Mismatches 93	Indels 18	Gaps 3
Qy	547	ACTGAGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAAGAAATCCNNAAGCCTTANCTGG	606	
Db	6498	AGTGAGAGACAGACTAGCTGGATGTCGAGGCCGACTAATATCCCTTAGCCTTAGCTGCTGG	6439	
Qy	607	GAAGGTGACCGCATCCATCTTTTAAACATATGGGGCTTGGCAACTTAGCTCACACCCGACCAAT	666	
Db	6438	GAAGGTGACCGCATCCACTTTTAAACACAGGGGCATGCAATTTAGCTCACACCCAAACCAAT	6379	
Qy	667	C-----AGAGAGCTCACTAAAAATGCTTAATCAGGCCAAACACAGGAGTGAAGCAATA	717	
Db	6378	CAGGTAGTAAGAGAGCTCACTAAAGTGCTAGTTAGGCCAAAAATAGGAAGTAAAGAAATA	6319	
Qy	718	GCCAAATCATCTATTTCCTTGAGAGACACAGCGGGAAGGACAAGGATTTGGGATATAAACTCAG	777	
Db	6318	GCCAAATCATCTATCACCTGAGAGACACAGGGGAAGGACAATGATCTGGATAGAAACCCAG	6259	
Qy	778	GCATTCAAGCCAGCAACAGCAACCCCCCTTTGGGGTCCCTCCCATTTGATATGGGAGCTCTGT	837	
Db	6258	GCATTCCAGCCAGCAACGGGTACCCAATTGGGGTCCCTCCCGTTGATATGGGAACTCTGT	6199	
Qy	838	TTTTCACCTCTATTTCACCTCTATTAAATCATGCAACTCTTCTGGTCCGTGTTTTTTA	897	
Db	6198	TTTTCACCTCTATTAAACTTTGMAACTGCA-----CACTTCTGGTCCGTGCTCTGTTA	6147	
Qy	898	TGGCTCAAGCTGAGCTTTTGTTCGCCATCCACCACTGCTGTTTGGCCACCGTCACAGACCC	957	
Db	6146	CGGCTTCAGCTGAGCTTTTGGCTCGCCGTCACCACTGCTGTTTCCGCCCGTCGAGACCC	6087	
Qy	958	GCTGCTGACCTTCATCCCTTTTGGATCCAGCAGAGTGTCCACTGTGCTCTGTATCCAGGGA	1017	
Db	6086	GCOCGTCACTTCCA-CCCTCCGGAATCCGGCAGGGGTGCCACTGTGCTCGTGAATCCAGGGA	6028	
Qy	1018	GGTACCATTGGCCACTCCCGATCAGGCTAAGGCTTGCCATTGTTCTCTGCATGGCTAAGT	1077	
Db	6027	GACACCATTTGCCGCTCCCAATCGGGCTGAGGCTCGCCATTGTTCTGCACAGCTAACT	5968	
Qy	1078	GCCTGGGTTTGTCTCTAATAGAACTGAAACATGTGTTCACTGGGTTCCATGGTTCTCTTCAAT	1137	
Db	5967	GCCTGGGTTTCACTCTGATCGAGCTGAAACATGTGTCGTGGGTTCCAGGGTTCTCTCCGT	5908	
Qy	1138	GACCCAGGGCTTCTAAATAGAGCTATACACTCACCGCATGCGCCAAAGATTTCCATTCTCTTG	1197	
Db	5907	GACCCAGGGCTTCTAAATAGAGCTATAAACACTCACCGCAGGGCCAGAGATTCATTCTCTTG	5848	
Qy	1198	GTATCTGTGAGGCCAAGAAACCCCAAGGTACAGAAANGTGAGGCTTTGCCACCAATTTGGGAAG	1257	
Db	5847	GAATCCGTGAGGCCAAGAAACCCCAAGGTACAGAAACACAGGCTTGTGCTCTCTTGGAG	5788	
Qy	1258	TGGCC 1262		
Db	5787	TGGCC 5783		

RESULT 29

```

RES001 29
US-09-731-231A-3/c
; Sequence 3, Application US/09731231A
; Patent No. US20020082189A1
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001007
; CURRENT APPLICATION NUMBER: US/09/731,231A
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Fast-SEQ for Windows Version 4.0

```

: SEO ID NO 3

```
?
? LENGTH: 326014
? TYPE: DNA
? ORGANISM: Human
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)-(326014)
? OTHER INFORMATION: n = A,T,C or G
US-09-731-231A-3
```

Query Match	37.8%;	Score 502;	DB 9;	Length 326014;
Best Local Similarity	88.7%;	Pred. No. 5.7e-144;		
Matches 557;	Conservative	0;	Mismatches 62;	Indels 9; Gaps 1;
Qy	257	CCCTGTATCTTCAACTTCCTTGTTAAAGTTTGTCTCTTCCAGAAATCGAAGCTGTAAAGCTA	316	
Db	170822	CCCTGTATCTTTNACTCTCTTGTTAAAGTTTGTCTCTTCCAAAATTGAGCTCTAAACTA	170763	
Qy	317	CAAAATAGTTCTTCAAAATGGAAACCCCAAGATGCAAGTCCATGACTTAAAAATCTACCGTGGACCC	376	
Db	170762	CAAAATGTTCTTCAAAATGGAGCCCAAGCGCAGTCCATGACTTAAGATCTACCGCAGACCC	170703	
Qy	377	CTGGA CGGCGCTGCTAGACTATGCTCTGATGTTAAATGACATTTGAAGTCAACCCCTCCCGAG	436	
Db	170702	CTGGA CGGCGCTGTAGCCCATGCTCCGATGTTAAATGACATCGAAGGCACCCCTCCAGAG	170643	
Qy	437	GAATCTCAACTGCACAACCCCTACTACACTCCAATTCAGTAGGAGCAGTTAGACGAGT	496	
Db	170642	GAATCTCAACTGCTATAACCCCCACTACGCCCAATTCAGCAGGAAAGCAGTTAGAGCGAGT	170583	
Qy	497	TGTCAGCAACCTCCCCAAACAGTACTTGGGTTTCTGTTGAGAGGTTGGACTTGAGAGAC	556	
Db	170582	CATCGGCCATCTCCCCAAACAGCATTTGGGTTTCTGTTGAGAGGGGTACTCAGAGAC	170523	
Qy	557	AGGACTAGCTGGGATTTTCTAGGCTGACTAAGAAATCCNAAAGCCTANCTGSGAAAGGTGACC	616	
Db	170522	AGGACTAGCTGGATTTTCTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGSGAAGGTGACT	170463	
Qy	617	GCATCATCTTTAAACATGGGGCTTGCAACTTAGCTCAACCCCGACCAATC-----	667	
Db	170462	GCCTTCTACCTTTAAACCCGGGGCTTGCAACTTAGCTCAACCTTGACCAATCAGGTAGGAA	170403	
Qy	668	AGAGAGCTCACTAAAATGCTTAATCAGGCAAAAACAGAGGTAAGCAATAGCCAAATCATC	727	
Db	170402	AGAGAGCTCACTAAAATGTTAACTAGGCTAAAACAGAGGTAAGAAAATAGCCAAATCATC	170343	
Qy	728	TATTGGCTCAGAGACACAGCGGGAAGGACAGGATTGGGATATAAACTCAGGCATTTCAAGC	787	
Db	170342	TATCGCTGAGAACACAGTGGGNGGACAAATGATTTGGGTTATAAACCCAGGCATTTGGAGC	170283	
Qy	788	CAGCAACAGCAACCCCTTTGGGTCCCTCCCATTTGATGGGAGCTCTGTTTTCACTCTA	847	
Db	170282	CAGCAATGGCTACCAATTTTGGGTCCCTCCCTTTGTATGGGAGCTCTGTTTTCACTCTA	170223	
Qy	848	TTTCACTCTTATTAATCATGCAACTGCA	875	
Db	170222	TTTCACTCTTATTAATCTTGCACAGCA	170195	

RESULT 30

```

RES001.30
US-10-751-985-3/C
; Sequence 3, Application US/10751985
; Publication No. US20040126861A1
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001007CON
; CURRENT APPLICATION NUMBER: US/10/751,985
; CURRENT FILING DATE: 2004-01-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 3
; LENGTH: 326014
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)....(326014)
; OTHER INFORMATION: n = A,T,C or G
; US-10-751-985-3

Query Match          37.8%; Score 502; DB 17; Length 326014;
Best Local Similarity 89.7%; Pred. No. 5.7e-144;
Matches 557; Conservative 0; Mismatches 62; Indels 9; Gaps 1;

QY 257 CCCTGTAATCTTCAACTTCCTTGGTTAAAGTTGTCTCTCCAGAAATTGAAGCTGTAAAGCTA 316
DB 170822 CCCTGTATCTTTTAACTCTCTTGTAAAGTTGTCTCTCCAAAATTGAAGCTCTAAACTA 170763

QY 317 CAATACTGTTCTTCAAAATGGAAACCCAGATGCAGTCCTAATCTAAAAATCTACCGTGGACCC 376
DB 170762 CAATAATGTTCTTCAAAATGGAGCCCAAGCGAGTCCATCATAGATCTACGCGAGACCC 170703

QY 377 CTGACCGCGCTGCTAGACTATGCTCTGAATGTTAATGACTGAAGTCAACCCCTCCCGAG 436
DB 170702 CTGACCGCGCTGCTAGCCCATGCTCCGATGTTAATGACATCGAAGGCACCCCTCCAGAG 170643

QY 437 GAAATCTCAACTGCACAAACCCCTACTACACTCCAAATTCAGTAGGNAGCAGTTAGNAGCAGT 496
DB 170642 GAAATCTCAACTGCTAATAACCCCACTAGCCCCCAATTGAGAGGAAGCAGTTAGAGCAGT 170583

QY 497 TGTGAGCGCAACTCTCCCAACAGTACTTGGGTTTTCTCTTGAGAGGGTGGACTGAGAGAC 556
DB 170582 CATCGGCATCCTCCCAACAGCAGCTTGGGTTTTCTCTTGAGAGGGGTACTCAGAGAC 170523

QY 557 AGGACTAGCTGGATTTCCTTAGCTTGACTAAGAAATCCCAAGCTTACTCTGGGAAGTGAACC 616
DB 170522 AGGACTAGCTGGATTTCCTTAGCTTGACTAAGAAATCCCTAAGCTAGCTTGGGAAGTGAAC 170463

QY 617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACCCCGACCAATC----- 667
DB 170462 GCTTTACCTTTTAAACCGGGGGCTTGCAACTTAGCTCACCTGACCAATCAGGTAGGAA 170403

QY 668 AGAGAGCTCACTAAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCAATCATC 727
DB 170402 AGAGAGCTCACTAAAATGCTAATCAGGCTAAAACAGGAGGTAAAGAAATAGCCAATCATC 170343

QY 728 TATTGCTTGAGAGCACACGGGGAAGGACAAGGATTTGGATATAACTCAGGCATTCAGC 787
DB 170342 TATCGCTTGAGAACACATGTTGGGAGGACAATGATTGGGTTTATAAACCCAGGCATTTGGAGC 170283

QY 788 CAGCAACAGCAACCCCTTTTGGGTCCCTCCCATTTGATGGAGCTCTGTTTTCACCTCTA 847
DB 170282 CAGCAATGGCTACCATTTTGGGTCCCTCCCTTTGATGGAGCTCTGTTTTCACCTCTA 170223

QY 848 TTTCACTCTATTAATCATGCAACTGCA 875
DB 170222 TTTCACTCTATTAATCTTTGCAACAGCA 170195

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2005, 12:01:19 ; Search time 3958.84 Seconds
(without alignments)
12232.983 Million cell updates/sec

Title: US-09-319-156b-12
Perfect score: 1329
Sequence: 1 tcaaatcgaagcttag.....gcaagatccccagtaaca 1329

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875.seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	706.8	53.2	2748	3	CR605851 full-leng
2	702.8	52.9	2749	3	CR617248 full-leng
3	702	52.8	2500	3	CR622175 full-leng
4	699	52.6	2748	3	CR625046 full-leng
5	684.2	51.5	1071	5	EX35066 BX365066
6	678.6	51.1	2716	3	CR613169 full-leng
7	669.4	50.4	998	5	EX337769 BX337769
8	661.2	49.8	1058	5	EX378303 BX378303
9	652.2	49.1	1019	5	EX439636 BX439636
10	610	45.9	966	5	EX380176 BX380176
11	608.2	45.8	758	5	EX357208 BX357208
12	593.4	44.7	771	4	BI087886 BX087886
13	592	44.5	4204	3	BC035153 Homo sapi
14	561	42.2	719	7	CN272394 170006000
15	544	40.9	900	5	EX459153 BX459153
16	500.6	37.7	1002	5	EX953822 BX953822
17	480.8	36.2	679	9	AG113694 Pan trogl
18	480.8	36.2	999	4	BM918330 AGENCOURT
19	475.2	35.8	683	9	AG134524 Pan trogl
20	474	35.7	669	9	AG049953 Pan trogl
21	473	35.6	681	9	AG116971 Pan trogl
22	469.6	35.3	710	9	AG096341 Pan trogl
23	468.2	35.2	1160	3	BC030968 Homo sapi
24	465.2	35.0	893	5	BQ437925 AGENCOURT

ALIGNMENTS

RESULT 1
CR605851
LOCUS full-length cDNA clone CS0DE012VJ24 of Placenta of Homo sapiens
DEFINITION (human).
ACCESSION CR605851
VERSION CR605851.1 GI:50486658
KEYWORDS HTC; CNSLT cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2748)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue
2 (bases 1 to 2748)
REFERENCE 2 (bases 1 to 2748)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)

COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.

FEATURES
Location/Qualifiers
1..2748
/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DE012VJ24"

/tissue_type="Placenta"

/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 53.2%; Score 706.8; DB 3; Length 2748;
Best Local Similarity 90.5%; Pred. No. 1.1e-202;
Matches 769; Conservative 0; Mismatches 69; Indels 12; Gaps 1;

Qy 1 TCAAAATCGAAGAGCTTTAGACTTGTCTTACCGCAAAAGAGGGGACCTGTTATTTTT 60
|||||

Db 1911 TCAAAATCGAAGAGCTTTAGACTTGTCTTACCGCTGAAGAGGGGACCTGTTATTTTT 1970
|||||

QY 61 AGGGAAGAATGCTGTAGTATGTTAATCAATCTGGAATCATTAATCTGAGAAGTTAAAGA 120
 Db |||||
 1971 AGGGAAGAATGCTGTATTTATTTAATCAATCCGGAATCGTCACTGAGAAGTTAAAGA 2030
 QY 121 AATTGAGATCGAATATAATGATAGACAGAGAGACCTTCAAAACACTGACCTGGGGCCT 180
 Db |||||
 2031 AATTGAGATCGAATATACAACGATAGACAGAGAGACCTTCAAAACACTGACCTGGGGCCT 2090
 QY 181 CCTCAGCAATGGAATGCCCTGACTCTCCCTCTCTAGGACCTCTAGCAGCTATAATTT 240
 Db |||||
 2091 CCTCAGCAATGGAATGCCCTGACTCTCCCTCTCTAGGACCTCTAGCAGCTATAATTT 2150
 QY 241 TTTACTCCTCTTTGGACCCCTGTATCTTCAACTCTCTTTGTTAAGTTTCTCTCTCCAGAA 300
 Db |||||
 2151 GCTACTCCTCTTTGGACCCCTGTATCTTTAACTCTCTTTGTTAAGTTTCTCTCTCCAGAA 2210
 QY 301 TGAAGCTGTAAGCTACAATAGTTCTTCAATGGAACCCAGATGAGTCCATGACTAA 360
 Db |||||
 2211 CGAAGCTGTAAACTA-----CAAATGGAGCCCAAGATGAGTCCCAAGACTAA 2258
 QY 361 AATCTACCGTGGACCCCTGGACCGGCTGTAGACTATGCTCTGATGTTAATGACATTGA 420
 Db |||||
 2259 GATCTACCGAGACCCCTGGACCGGCTGTAGCCGAGATCTGATGTTAATGACATCAA 2318
 QY 421 AGTCACCCCTCCGAGGAAATCTCAACTGCACCAACCCCTACTACCTCCAATTCAGTAGG 480
 Db |||||
 2319 AGGCACCCCTCCTGAGGAATCTCAGCTGCACAACTCTACTAGCCGCCAATTCAGCAGG 2378
 QY 481 AAGCAGTTAGACGAGTTGTGAGCCAACTCCCAACAGTACTTGGGTTTCTGTTGAGA 540
 Db |||||
 2379 AAGCAGTTAGACGCGTCTGCGGCCAACCTCCCAACAGCACTTAGGTTTCTGTTGAGA 2438
 QY 541 GGGTGGACTGAGACAGGACTAGCTGATTTCTAGGCTGACTAAGAATCCCAAGCCT 600
 Db |||||
 2439 TGGGGACTGAGACAGGACTAGTGGATTTCTAGGCTGACTAAGAATCCCAAGCCT 2498
 QY 601 ANCTGGAAGGTGACCGCATCCATCTTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
 Db |||||
 2499 AGCTGGGAAGGTGACCATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2558
 QY 661 ACCAATCAGAGCTCACTAAATGCTTAATCAGGCAAAACAGGAGGTAAAGCAATAGCC 720
 Db |||||
 2559 ACCAATCAGAGCTCACTAAATGCTTAATCAGGCAAAACAGGAGGTAAAGCAATAGCC 2618
 QY 721 AATCATCTATTGCTGAGAGACAGCGGGAAGGACAGGATTCGGATATAAATCTCAGGCA 780
 Db |||||
 2619 AATCATCTATTGCTGAGAGACAGCGAGGAGGACATGATCGGATATAAACCAGTC 2678
 QY 781 TTAAGCAGCAACAGCAACCCCTTTGGTCCCTCCCAATGATGGGAGCTCTGTTT 840
 Db |||||
 2679 TTAAGCAGCAACAGCAACCCCTTTGGTCCCTCCCAATGATGGGAGCTCTGTTT 2738
 QY 841 CACTCTATTT 850
 Db |||||
 2739 CATGCTATTT 2748

RESULT 2
 LOCUS CR617248
 DEFINITION full-length cDNA clone CS0D1022YJ18 of Placenta Cot 25-normalized
 of Homo sapiens (human)
 ACCESSION CR617248
 VERSION CR617248.1 GI:50498055
 KEYWORDS HTC; CNSLT_cDNA
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 2749)
 Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization

JOURNAL
REMARK

Unpublished
 Contact : Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Paraday Avenue
 2 (bases 1 to 2749)

REFERENCE
AUTHORS
TITLE
JOURNAL

Genoscope.
 Direct Submission
 Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

COMMENT
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen.

FEATURES

source

1..2749
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0D1022YJ18"
 /tissue_type="Placenta Cot 25-normalized"
 /plasmid="pCMVSPORT_6"

ORIGIN

Query Match 52.9%; Score 702.8; DB 3; Length 2749;
 Best Local Similarity 90.4%; Pred. No. 1.8e-201;
 Matches 765; Conservative 0; Mismatches 69; Indels 12; Gaps 1;

QY 1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCCGCAAAAGAGGGGGAACCTGTTTATTTT 60
 Db 1916 TCAAAATCGAAGAGCTTTAGACTTGCTAACCCGCTGAAAGAGGGGGAACCTGTTTATTTT 1975
 QY 61 AGGGGAAGAATCGCTGTAGTATGTTAATCAATCTGMACTTACTGAGAAAGTTAAAGA 120
 Db 1976 AGGGGAAGAATCGCTGTATTTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2035
 QY 121 AATTTGAGATCGAATATAATGATAGACAGAGAGACCTTCAAAACACTGACCCCTGGGGCCT 180
 Db 2036 AATTCGAGATCGAATACACGATAGACAGAGAGCTTCGAAACACTTGGACCCCTGGGGCCT 2095
 QY 181 CCTCAGCAATGGAATGCCCTGGAATCTCCCTCTTTAGGACCTCTAGCAGCTATAATTT 240
 Db 2096 CCTCAGCAATGGAATGCCCTGGAATCTCCCTCTTTAGGACCTCTAGCAGCTATAATTT 2155
 QY 241 TTTACTCCTCTTTGGACCCCTGTATCTCAACTCTCTGTTAAGTTTCTCTCTCCAGAA 300
 Db 2156 GCTACTCTCTTTGGACCCCTGTATCTTTAACTCTTGTGTTAAGTTTCTCTCTCCAGAA 2215
 QY 301 TGAAGCTGTAAAGCTACAAATAGTTTCTTCAATGGAACCCCAAGATGCACTCCATGACTAA 360
 Db 2216 CGAAGCTGTAAACTA-----CAAATGGAGCCCAAGATGCACTCCCAAGACTAA 2263
 QY 361 AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGATGTTAATGACATTGA 420
 Db 2264 GATCTACCGCAGACCCCTGGACCGGCTGCTAGCCCAAGATCTGATGTTAATGACATCA 2323
 QY 421 AGTCACCCCTCCGAGGAAATCTCAACTGCACCAACCCCTACTACACTCCAAATTCAGTAGG 480
 Db 2324 AGGCACCCCTCTGAGGAAATCTCAGCTGCACCAACTCTATACGCCCCCAATTCAGCAGG 2383
 QY 481 AAGCAGTTAGACGAGTTGTGAGCCAACTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 540
 Db 2384 AAGCAGTTAGACGAGTTGTGAGCCAACTCCCAACAGCAGCTTAGGTTTCTCTGTTGAGA 2443
 QY 541 GGGTGGACTGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAAGATCCCAAGCCT 600
 Db 2444 TGGGGGACTGAGACAGGACTAGCTGGATTTCTTAGGCTGACTAAGATCCCTAAGCCT 2503
 QY 601 ANCTGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
 Db 2504 AGCTGGGAAGGTGACCATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 2563

```
QY 661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAAACAGAGGTAAAGCAATAGCC 720
    |||
    |||
    |||
Db 2564 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAAACAGAGGTAAAGCAATAGCC 2623
    |||
    |||
    |||
QY 721 AATCATCTATTGCTGAGAGCAGAGCGGGAGGAGCAAGGATTCGGATATAAATCAGGCA 780
    |||
    |||
    |||
Db 2624 AATCATCTATTGCTGAGAGCAGAGCGGGAGGAGCAAGGATTCGGATATAAATCAGGCA 2683
    |||
    |||
    |||
QY 781 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCTCCCTATGATGGGAGCTCTGTTT 840
    |||
    |||
    |||
Db 2684 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCTCCCTATGATGGGAGCTCTGTTT 2743
    |||
    |||
    |||
QY 841 CACTCT 846
    |||
    |||
    |||
Db 2744 CATGCT 2749
    |||
    |||
    |||

RESULT 3
CR622175
LOCUS 2500 bp mRNA linear HTC 21-JUL-2004
DEFINITION full-length cDNA clone CS0D1051YM13 of Placenta Cot 25-normalized
of Homo sapiens (human).
ACCESSION CR622175
VERSION 1 GI:50502982
KEYWORDS HTC; CNSLT_cDNA.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
JOURNAL Full-length cDNA libraries and normalization
REMARK Unpublished
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue
Genoscope.
2 (bases 1 to 2500)
REFERENCE Direct Submission
AUTHORS Genoscope.
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source Location/Qualifiers
1. .2500
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1051YM13"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 52.8%; Score 702; DB 3; Length 2500;
Best Local Similarity 90.6%; Pred. No. 3.le-201;
Matches 763; Conservative 0; Mismatches 67; Indels 12; Gaps 1;

QY 1 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCCAAAGAGGGGAACTGTTATTTT 60
Db 1669 TCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAGAGGGGAACTGTTATTTT 1728

QY 61 AGGGGAAGATGCTGTTAGTATGTTAATCAATCTGGAAATCATTAATGAGAAAGTTAAGA 120
Db 1729 AGGGGAAGATGCTGTTATTTATGTTAATCAATCGGAATCGTCACTGAGAAAGTTAAGA 1788
    |||
    |||
    |||
QY 121 AATTGAGATCGAATATAAGTAGCAGAGGAGCTTCAAAACACTGCACCTGGGGCCT 180
    |||
    |||
    |||
Db 1789 AATTGAGATCGAATACAACGTAGCAGAGGAGCTTCGAAACACTGGACCTGGGGCCT 1848
```

```
QY 181 CCTCAGCCAATGGATGCCCTTGGACTCTCCCTTTCTTTAGGACCTCTAGCAGCTATAATATT 240
    |||
    |||
    |||
Db 1849 CCTCAGCCAATGGATGCCCTTGGATTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATATT 1908
    |||
    |||
    |||
QY 241 TTTACTCTCTTTTGGACCTCTGATATCTTCAACTTCCTTTAAAGTTTGTCTCTTCCAGAAAT 300
    |||
    |||
    |||
Db 1909 GCTACTCTCTTTTGGACCTCTGATATCTTAACTTCCTTTAACTTTGTCTCTTCCAGAAAT 1968
    |||
    |||
    |||
QY 301 TGAAGCTGTAAAGCTACAATAATAGTTCTTCAAAATGGAAACCCAGATGCAGTCCATGACTAA 360
    |||
    |||
    |||
Db 1969 CGAAGCTGTAAACTA-----CAAAATGGAGCCCAAGATGCAGTCCAAGACTAA 2016
    |||
    |||
    |||
QY 361 AATCTACCGTGGACCCCTTGGACCGGCTGCTAGACTATGCTCTGATGTTTAAATGACATTTGA 420
    |||
    |||
    |||
Db 2017 GATCTACCGCAGACCCCTTGGACCGGCTGCTAGCCCAACGATCTGATGTTTAAATGACATCA 2076
    |||
    |||
    |||
QY 421 AGTCACCCCTCCGAGGAAATCTCAACTGCACAAACCCCTACTACACTCCAAATTCAGTAGG 480
    |||
    |||
    |||
Db 2077 AGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACTCTACTACGCCCAATTCAGCAGG 2136
    |||
    |||
    |||
QY 481 AAGCAGTTAGAGCAGTTGTTCAGCCAACTCCCAACAGTACTTGGCTTTTCTCTGTTGAGA 540
    |||
    |||
    |||
Db 2137 AAGCAGTTAGAGCAGTTGTTCGCGCAACCTCCCAACAGCAGCTTAGGTTTCTCTGTTGAGA 2196
    |||
    |||
    |||
QY 541 GGGTGACTTGAGAGCAGGACTAGCTGGATTTTCTTAGGCTGACTAAGAAATCCCAAGACCT 600
    |||
    |||
    |||
Db 2197 TGGGGACTTGAGAGCAGGACTAGCTAGCTGGATTTTCTTAGGCTGACTAAGAAATCCCAAGCCT 2256
    |||
    |||
    |||
QY 601 ANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCG 660
    |||
    |||
    |||
Db 2257 AGCTGGGAAGGTGACCAACATCCACCTTTAAACACAGGGGCTTGCAACTTAGCTCACACCTG 2316
    |||
    |||
    |||
QY 661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
    |||
    |||
    |||
Db 2317 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 2376
    |||
    |||
    |||
QY 721 AATCATCTATTGCTGAGAGCAGCGGGAAGGACAGAGGATTCGGATATAAATCTCAGGCA 780
    |||
    |||
    |||
Db 2377 AATCATCTATTGCTGAGAGCAGCGGGAAGGACAGGAGGAGCAATGATCGGATATAAATCAGGTC 2436
    |||
    |||
    |||
QY 781 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCTCCCTATGATGGGAGCTCTGTTT 840
    |||
    |||
    |||
Db 2437 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCTCCCTATGATGGGAGCTCTGTTT 2496
    |||
    |||
    |||
QY 841 CA 842
    |||
    |||
    |||
Db 2497 CA 2498
    |||
    |||
    |||

RESULT 4
CR625046
LOCUS 2748 bp mRNA linear HTC 21-JUL-2004
DEFINITION full-length cDNA clone CS0D1044YK06 of Placenta Cot 25-normalized
of Homo sapiens (human).
ACCESSION CR625046
VERSION 1 GI:50505853
KEYWORDS HTC; CNSLT_cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
JOURNAL Full-length cDNA libraries and normalization
REMARK Unpublished
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue
Genoscope.
2 (bases 1 to 2748)
REFERENCE Direct Submission
AUTHORS Genoscope.
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
```

```
- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

FEATURES
    source          Location/Qualifiers
    ORIGIN
        Query Match          52.6%; Score 699; DB 3; Length 2748;
        Best Local Similarity 90.6%; Pred. No. 2.6e-200;
        Matches 760; Conservative 0; Mismatches 67; Indels 12; Gaps 1;
        QY 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCAAAAGAGGGGAACTGTTTATTTT 60
        DB 1922 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCTGAAGAGGGGAACTGTTTATTTT 1981
        QY 61 AGGGGAAGAAATGCTGTGTATGTTAATCAATCTGGAATCAATCTAGAGAAAGTTAAAGA 120
        DB 1982 AGGGGAAGAAATGCTGTGTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2041
        QY 121 AATTGAGATCGAATATAATGTAGACGAGGAGCTTCAAAACACTGCACCCCTGGGCGCT 180
        DB 2042 AATTGAGATCGAATATAACGTAGACGAGGAGCTTCAAAACACTGCACCCCTGGGCGCT 2101
        QY 181 CCTCAGCAATGATGATGCTTGGACTCTCCCTCTTCTAGGACCTCTAGCAGCTATAATATT 240
        DB 2102 CCTCAGCAATGATGATGCTTGGACTCTCCCTCTTCTAGGACCTCTAGCAGCTATAATATT 2161
        QY 241 TTACTCTCTCTTTGGACCCCTGTATCTTCAACTCTTCTGTTTAAAGTTGTCTCTTCCAGAAAT 300
        DB 2162 GCTACTCTCTCTTTGGACCCCTGTATCTTAACTCTCTTGTAACTTGTCTCTTCCAGAAAT 2221
        QY 301 TGAAGCTGTAAGCTACAAATAGTTCTTCAAAATGGAACCCCAAGATGAGTCCATGACTAA 360
        DB 2222 CGAAGCTGTAAACTA-----CAAAATGGAAGCCCAAGATGAGTCCCAAGACTAA 2269
        QY 361 AATCTACCGTGGACCCCTGACCGGCTGCTAGACTATGCTCTGATGTTTAAATGACATTTGA 420
        DB 2270 GATCTACCGGAGACCCCTGACCGGCTGCTAGCCACGATCTGATGTTAATGACATCAA 2329
        QY 421 AGTCACCCCTCCCGAGGAATCTCAACTGCACAAACCCCTTACTACACTCCCAATTCAGTAGG 480
        DB 2330 AGGCACCCCTCTCTGAGGAATCTCAGCTGCACAACTCTACTACGCCCAATTCAGCAGG 2389
        QY 481 AAGCAGTTAGACAGATGTTGAGCAACCTCCCAACAGTACTTGGGTTTCTCTGTTGAGA 540
        DB 2390 AAGCAGTTAGAGCGGTGCTCGGCCAACCCTCCCAACAGCACTTTAGGTTTCTCTGTTGAGA 2449
        QY 541 GGGTGGACTCAGAGACGAGTACTAGCTGATTTCTTAGGCTGACTAAGAATCCCAAGAGCCT 600
        DB 2450 TGGGGACTCAGAGACGAGTACTAGCTGATTTCTTAGGCTGACTAAGAATCCCAAGAGCCT 2509
        QY 601 ANCTGGGAAGGTACCGCACTCCATCTTTAAACATGGGGCTTGCACCTAGCTCAGCAGCCG 660
        DB 2510 AGCTGGGAAGGTACCCACATCCACCTTTAAACACGCGGCTTGCACTAGCTCAGCAGCTG 2569
        QY 661 ACCAATCAGAGCTCCTAAATGCTAATCAGGCAAAACAGGAGGTAAAGCAATAGCC 720
        DB 2570 ACCAATCAGAGCTCCTAAATGCTAATTAGGCAAAACAGGAGGTAAAGCAATAGCC 2629
        QY 721 AATCATCTATTGCTCAGAGACAGCGGGAAGGACAGGATTCGGGATATAAATCTCAGGCA 780
        DB 2630 AATCATCTATTGCTCAGAGACAGCGGGAAGGACAGTATCGGGATATAAATCCCAAGTC 2689
        QY 781 TTCAAGCCAGCAACAGCAACCCCTTTGGGTCCCTCTCCCATTTGATGGGAGCTCTGTTT 839
```

```
Db 2690 TTCGAGCCGCAACGCAACCCCTTTGGTCCCTCTTGTATGGAGCTCTGTTT 2748

RESULT 5
LOCUS      BX365066/c
DEFINITION BX365066 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI046YA18 3-PRIME, mRNA sequence.
ACCESSION  BX365066
VERSION    BX365066.2 GI:46304105
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1071)
AUTHORS   Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   On May 5, 2003 this sequence version replaced gi:30374869.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            4215.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS1A1012ZC10NP1&c=4215.r.

FEATURES
    source          Location/Qualifiers
    1..1071
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="CS0DI046YA18"
        /tissue_type="PLACENTA COT 25-NORMALIZED"
        /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
        /note="1st strand cDNA was primed with a NotI-oligo(dT)
        primer. Five prime end enriched, double-strand cDNA was
        digested with Not I and cloned into the Not I and EcoR V
        sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match          51.5%; Score 684.2; DB 5; Length 1071;
Best Local Similarity 89.9%; Pred. No. 6.1e-196;
Matches 755; Conservative 3; Mismatches 69; Indels 13; Gaps 2;
QY 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCAAAAGAGGGGAACTGTTTATTTT 60
DB 827 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCTGAAGAGGGGAACTGTTTATTTT 768
QY 61 AGGGGAAGAAATGCTGTGTATGTTAATCAATCTGGAATCATTTCTGAGAAAGTTAAAGA 120
DB 767 AGGGGAAGAAATGCTGTGTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 708
QY 121 AATTGAGATCGAATATAATGTAGACGAGGAGCTTCAAAACACTGCAACCTCGGGCGCT 180
DB 707 AATTGAGATCGAATATAACAGTGAAGAGGAGCTTCGAAACACTGACCCCTGGGCGCT 648
QY 181 CCTCAGCAATGATGATGCTTGGACTCTCCCTCTTCTAGGACCTCTAGCAGCTATAATATT 240
DB 647 CCTCAGCAATGATGATGCTTGGACTCTCCCTCTTCTAGGACCTCTAGCAGCTATAATATT 588
QY 241 TTACTCTCTCTTTGGACCCCTGTATCTTCAACTCTTCTGTTTAAAGTTGTCTCTTCCAGAAAT 300
DB 587 GCTACTCTCTCTTTGGACCCCTGTATCTTAACTCTCTTGTAACTTGTCTCTTCCAGAAAT 528
QY 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAATGGAACCCCAAGATGCACTGATGACTAA 360
```

```
Db 527 CGAGCTGTAAACTA-----CAAATGGAGCCCAAGATGCAGTCCAGACTAA 480
Qy 361 AATCTACCGTGGACCCCTGCAGCGGCTGTAGACTATGCTGTGATGTTAATGACATTTGA 420
Db 479 GATCTACCGCAGACCCCTGCAGCGGCTGTAGCCACGATCTGTGTTAATGACATCAA 420
Qy 421 AGTCACCCCTCCGAGGAATCTCAACTGACACACCCCTACTACACTCCAAATTCAGTAGG 480
Db 419 AGGACACCCCTCCGAGGAATCTCAGCTGCACAACTCTACTACGCCCAATTCAGGAGG 360
Qy 481 AAGCAGTTAGAGCAGTGTGTCAGCCAACTCCCAACAGATCTTTGGGTTTTCCTGTTGAGA 540
Db 359 AAGCAGTTAGAGCGGTCTGCGCCAACTCCCAACAGACTTTAGGTTTTCCTGTTGAGA 300
Qy 541 GGTGAGCTGAGAGACAGGACTAGCTGAGATTTCTTAGGCTGACTAAGAAATCCNAAGCCT 600
Db 299 TGGGGGACTGAGAGACAGGACTAGCTGAGATTTCTTAGGCTGACTAAGAAATCCCTAAGCCT 240
Qy 601 ANCTGGGAGGTGACCGCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCGG 660
Db 239 AGCTGGGAGGTGACCACTCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 180
Qy 661 ACCAATCAGAGAGTCACTAAATGCTTAATCAGGCAAAAAACAGGAGTAAAGCAATAGCC 720
Db 179 ACCAATCAGAGAGTCACTAAATGCTTAATCAGGCAAAAAACAGGAGTAAAGCAATAGCC 120
Qy 721 AATCATCTATTGCTGAGAGCAGCGGAGGAGCAAGGATTCGGGATATAAACTCAGGCA 780
Db 119 AATCATCTATTGCTGAGAGCAGCGGAGGAGGACAAATGATCGGATATAAAACCAAGTC 60
Qy 781 TTCAAGCCAGCAACAGACCCCTTTGGTCCCTCCCAATGTTATGGGAGCTCTGTTTT 840
Db 59 TTGAGCGCGCAACGCGCAACCCCTTTGGTNCCTCCCTCC-TTGTATGCGASMTCKGTTTT 1

RESULT 6
CR613169
LOCUS
DEFINITION
full-length cDNA clone CS0D013Y120 of Placenta of Homo sapiens
(human).
ACCESSION
CR613169.1 GI:50493976
VERSION
HTC; CNSLT_CDNA.
KEYWORDS
Homo sapiens
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 2716)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue
Genoscope.
2 (bases 1 to 2716)
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1..2716
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D013Y120"
/tissue_type="Placenta"
/plasmid="pCMVSPORT_6"

Query Match 51.1%; Score 678.6; DB 3; Length 2716;
Best Local Similarity 90.5%; Pred. No. 4.1e-194;
Matches 739; Conservative 0; Mismatches 66; Indels 12; Gaps 1;

ORIGIN
Qy 1 TCAAAATCGAAGAGCTTTAGACTTGTCTAAACGCGCAAAAGAGGGGGGAACCTGTTTATTTT 60
Db 1912 TCAAAATCGAAGAGCTTTAGACTTGTCTAAACGCGTGAAGAGGGGGGAACCTGTTTATTTT 1971
Qy 61 AGGGGAAGAAATGCTGTTAGTATGTTAATCAATTCGAATCATTTACTGAGAAAGTTAAAGA 120
Db 1972 AGGGGAAGAAATGCTGTTATTTATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 2031
Qy 121 AATTGAGATCGAATAATAATGTAGACAGAGAGACCTTCAAAAACATGACACCTCGGGGCT 180
Db 2032 AATTGAGATCGAATAATAACGTAGACAGAGAGAGCTTCGAAAACATGACACCTCGGGGCT 2091
Qy 181 CCTCAGCCCAATGGATCCCTGGACTCTCCCTCTCTTGGACCTCTAGCAGCTATAATTT 240
Db 2092 CCTCAGCCCAATGGATCCCTGGATTCTCCCTTCTTAGGACCTCTAGCAGCTATAATTT 2151
Qy 241 TTACTCTCTTTGGAGCCCTGTATCTTCAACTTCTTGTGTTAAGTTTGTCTTCCAGAA 300
Db 2152 GCTACTCTCTTTGGAGCCCTGTATCTTAACTCTCTTGTGTTAATGTTCTTCCAGAA 2211
Qy 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAATGGAAACCCAGATGACGTCATGACTAA 360
Db 2212 CGAAGCTGTAAACTA-----CAAATGGAGCCCAAGATGCAGTCCAGACTAA 2259
Qy 361 AATCTACCGTGGACCCCTCGACCGGCTGTAGACTATGCTCTGATGTTAATGACATTTGA 420
Db 2260 GATCTACCGCAGACCCCTCGACCGGCTGTAGCCCAAGATCTGATGTTTAAATGACATCAA 2319
Qy 421 AGTCACCCCTCCGAGGAAATCTCAACTGCACAACCCCTACTACTACTTCCAAATTCAGTAGG 480
Db 2320 AGGCACCCCTCTCTGAGGAAATCTCAGCTGCACAACTCTACTACGCCCAATTCAGCAGG 2379
Qy 481 AAGCAGTTAGAGCAGTTGTGAGCCAACTCCCAACAGTACTTGGGTTTCTGTTGAGA 540
Db 2380 AAGCAGTTAGAGCGGTGTCGCGCAACCTCCCAACAGCAGTCTAGGTTTCTGTTGAGA 2439
Qy 541 GGTGAGCTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGATCCCNAGCCT 600
Db 2440 TGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAAATCCCTAAGCCT 2499
Qy 601 ANCTGGGAAAGGTGACCGCATCCATCTTTAAACATGGGCTTGCAACTTAGCTCACACCCG 660
Db 2500 AGCTGGGAAAGGTGACCACTCCACCTTTAAACAGGGGCTTGGCACTTAGCTCACACCTG 2559
Qy 661 ACCAATCAGAGAGCTCACTAAATGCTAATATCAGGCAAAAAACAGGAGTAAAGCAATAGCC 720
Db 2560 ACCAATCAGAGAGCTCACTAAATGCTAATATGAGCAAGAGAGAGTAAAGCAATAGCC 2619
Qy 721 AATCATCTATTGCTGAGAGCAGCGGAGGAGCAAGGATTCGGGATATAAACTCAGGCA 780
Db 2620 AATCATCTATTGCTGAGAGCAGCGGAGGAGCAATGATCGGATATAAAACCAAGTC 2679
Qy 781 TTCAAGCCAGCAACAGACCCCTTTGGTCCCTCCCTC 817
Db 2680 TTGAGCGCGCAACGCGCAACCCCTTTGGTCCCTCCCTC 2716

RESULT 7
BX337769/c
LOCUS
DEFINITION
BX337769 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1051YM13 3-PRIME, mRNA sequence.
ACCESSION
BX337769
VERSION
BX337769.2 GI:46272079
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
```



```

Db      247 AGCTGGGAGGTGACCAATCCACTTTAAACACGGGGCTTGCAACTTAGCTCACACTG 188
Qy      661 ACCAATCAGAGCTCACTAAATGCTAAATCAGGCAAAAACAGGAGGTAAAGCAATAGCC 720
Db      187 ACCAATCAGAGCTKACTAAATGCTAACTAGGCAAAKACAGGAGTWWKTAATAKNV 128
Qy      721 AATCATCTATTGCTGAGAGACAGCGGGAAGCAAGGATTGGGATATAAATCAGGCA 780
Db      127 RATTAATTATTGCTGAGAGACAGCAGGAGGACWATKATCGGATATAAAYCCAAGTC 68
Qy      781 TTCAGCCAGCAACAGCAACCCCTTTGGGTCCCTCCCAATTGATGGAGCTCTGTTTT 840
Db      67 TTCAGCCGCGCAACAGAAACCCYTTTT-TTTCYCYCTTTTGTGNTGGNDTTTTTTTN 9
Qy      841 CACTCTAT 848
Db      8 NTTATTT 1

RESULT 10
LOCUS   BX380176/c
DEFINITION BX380176 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1044YK06 3-PRIME, mRNA sequence.
ACCESSION BX380176
VERSION   BX380176.2 GI:46834414
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 966)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 8, 2003 this sequence version replaced gi:30460243.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
4215.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0D1044BF03NP1&c=4215.r.
FEATURES             source
Location/Qualifiers
1..966
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1044YK06"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 45.9%; Score 610; DB 5; Length 966;
Best Local Similarity 82.5%; Pred. No. 2.1e-173;
Matches 693; Conservative 24; Mismatches 110; Indels 13; Gaps 2;

Qy      1 TCAAAATCGAGACTTTAGACTTGCTACCGCCAAAGAGGGGGAACCTGTTATTTT 60
Db      830 TCAAAATCGAAGACTTTARACTTGCTAMCCGCTGAAGAGGGGGAACCTGTTATTTT 771
Qy      61 AGGGGAAGAAATGCTGTTAGTATCTTAATCAATCTGGAATCATTTACTGAGAAAGTTAAAGA 120
Db      770 AGGGGAAGAAATGCTGTTATGTTATCAATCCGGAATCGTCACTGAGAAAGTTAAGA 711

```

```

Qy      121 AATTTCAGATCGAATATAATGTAGACGAGGACCTTCAAAACACCTGCACCCCTGGGCGCT 180
Db      710 AATTTCAGATCGAATACAACGTAGACGAGGAGCTTCGAAACACCTGGGAGCCCTGGGCGCT 651
Qy      181 CCTCAGCCAAATGGATCCCTGGACTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATATT 240
Db      650 CCTCAGCCAAATGGATCCCTGGATTCTCCCTTTCTTAGGACCTCTAGCAGCTATAATATT 591
Qy      241 TTTACTCTCTTTTGGACCTGTATCTTCAACTCTCTTGTAGTTTGTCTCTTCAGAAAT 300
Db      590 GCTACTCTCTTTTGGACCTGTATCTTAACTCTCTTGTAACTTTGTCTCTTCCTCCAGAAAT 531
Qy      301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAAAATGGAACCCAGATCCAGTCCATAGCTAA 360
Db      530 CGAAGCTGTAAACTA-----CAATGGAGCCCAAGATCGAGTCCAGACTAG 483
Qy      361 AATCTACCTGTGACCCCTGGACCGGCTCTAGACTATGCTCTGTAGTTGTTAAATGACATTGA 420
Db      482 GDTCTACCGCGGCCCCCTGGACCGGCTCTGCTGCCCCCGMTCTGATGTTAAATGACATCAA 423
Qy      421 AGTCACCCCTCCGAGGAAATCTCACTGCACAAACCCCTACTACACTCCCAATTCAGTAGG 480
Db      422 AGCACCCCTCTCTGGGGCAATCTCAGCTGCACMCCCTCTACTACGCCCAATTCGGCRGG 363
Qy      481 AAGCAGTTAGAGCAGTTGTGAGCAACCTCCCAACAGTACTTTGGGTTTTCTCTGTTGAGA 540
Db      362 AAGCAGTTGGGGGGTCTGTCGGCAACCTCCCGACGCGACTTGGTTTTCTCTGTTGAGA 303
Qy      541 GGGTGAAGCTGAGAGACAGACTAGCTGGATTCTTAGGCTGACTAAGAAATCCCAAGCCT 600
Db      302 TGGGGGACTGAGAGACMGGACTGGCTGGATTCTTAGGCTGACTAAGAAATCCCTARGCCT 243
Qy      601 ANCTGGGAGGTGACCGCATCCATCTTTAAACATGGGCTTCAACTTAGCTCAGCACACCG 660
Db      242 RGTGGGAAGGTGACCATCTCCACTTTAAACMCGGGGCTTGCCACTTGGCTCACCCTCG 183
Qy      661 ACCAATCAGAGAGCTCACTAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGC- 719
Db      182 CCCAATCGGGGCTCACTAAATGCTAATGCTGCGGCCSSSSSSSSSSSSSSSSSSSSSSSS 123
Qy      720 CAATCATCTATTGCTGAGAGACAGCGGGAAGGACAGGATTGGGATATAAATCAGCAGC 779
Db      122 CMSYCTCTATTGCTGAGAGACAGCGGGAAGGCAATGATCGGGATATAAACCACCMGT 63
Qy      780 ATTCAAGCAGCAACAGCAACCCCTTTGGGTCCCTCCCAATTGATGGAGCTCTGTTT 839
Db      62 CTTGGGCGCGCCCGCCCCCTTTGGGTCCCTCTCTCTTCGTGNTGGAGCTCTGT 3

```

```

RESULT 11
LOCUS    BX357208/c
DEFINITION BX357208 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1022YJ18 3-PRIME, mRNA sequence.
ACCESSION BX357208
VERSION   BX357208.2 GI:46305959
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 758)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 5, 2003 this sequence version replaced gi:30376125.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned

```

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 4215.r

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0D1022D809NP1&c=4215.r.

FEATURES

source
1..758
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1022YJ18"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match 45.8%; Score 608.2; DB 5; Length 758;
Best Local Similarity 88.1%; Pred. No. 6.9e-173;
Matches 665; Conservative 9; Mismatches 69; Indels 12; Gaps 1;

QY 75 GTTAGTATGTTAATCAATCTGGAATCATTAATCTAGAAAGTTAAGAAATTTGAGATCGAA 134
DB 758 GTTATTATGTTAATCAATCGGAATCGTCACTGAGAAAGTTAAGAAATTTGAGATCGAA 699
QY 135 TATAATCTAGACGAGAGACCTTTCAAAACACTGCACCTGGGCGCTCTCAGCAATGGA 194
DB 698 TACACGTAGACGAGAGAGCTTCGAAACACTGACCTGGGCGCTCTCAGCAATGGA 639
QY 195 TGCCTGGACTCTCCCTCTTAGGACCTCTAGCAGCTATAATATTTTACTCTCTTTG 254
DB 638 TGCCTGGATTCTCCCTCTTAGGACCTCTAGCAGCTATAATATTTGCTCTCTTTG 579
QY 255 GACCTGTATCTTCAACTTCTTTAGTTTGTCTCTTCCAGAAATGAAGCTGTAAGC 314
DB 578 GACCTGTATCTTAACTCTTTTAACTTTGTCTCTTCCAGAAATGAAGCTGTAAGC 519
QY 315 TACAATAGTTCTTCAATAGGAACCCAGAGTGCAGTCCATGACTAAATCTACCGTGGAC 374
DB 518 TA-----CAATGGAGCCCAAGATGAGTCCCAAGATTAAGATCTACCGAGAC 471
QY 375 CCCTGGACCGGCTGTAGACTATGCTCTGATGTTAATGACATTAAGTCAACCCCTCCCG 434
DB 470 CCCTGGACCGGCTGTAGCCAGATCTGATGTTAATGACATCAAGGCAACCCCTCTG 411
QY 435 AGGAATCTCACTGACAAACCCCTACTACATCTCAATTCAGTAGGAAGCAGTTAGCA 494
DB 410 AGGAATCTCAGCTGCACAACTCTACTACGCCCAATTCAGCAGGAAGCAGTTAGAGC 351
QY 495 GTTGTGAGCAACCTCCCAACAGTACTTGGGTTTCTGTTGAGAGGTGGACTGAG 554
DB 350 GTGTCGGCCAACTCCCAACAGCACTTAGTTTCTCTTGGATGGGGGACTGAG 291
QY 555 ACAGGACTAGCTGATTTCTTAGCTGACTAAGAAATCCCAAGCTTANCTGGGAAGTGA 614
DB 290 ACAGGACTAGCTGATTTCTTAGCTGACTAAGAAATCCCAAGCTTANCTGGGAAGTGA 231
QY 615 CCACATCCATCTTTAAACATGGGGCTTGAACCTTAGCTCACACCCGACCAATCAGAGAG 674
DB 230 CCACATCCATCTTTAAACATGGGGCTTGAACCTTAGCTCACACCTGACCAATCAGAGAG 171
QY 675 TCACTAAATGCTAATCAGCAAAACAGAGGTAAAGCAATAGCCATCATCTATTGCC 734
DB 170 TCACTAAATGCTAATTAGCAAGAGCAGAGGTAAAGCAATAGCCATCATCTATTSCC 111
QY 735 TGAGAGCACAGCGGAGGACAGGATTGGGATATAAACTTCAGGCTTCAAGCCAGCAAC 794
DB 110 YSAGACACAGCAGGAGGACATGATCGGGATATAAACCCCMMSYCCCCCGCGGCCAC 51
QY 795 AGCAACCCCTTTGGGTCCCTCCCATTTGATGGG 829

Db 50 GGCMACCCCCYCTGGTCCCTCCCTTCTTGTATGGG 16

RESULT 12

LOCUS BI087886
DEFINITION 771 bp mRNA linear EST 20-JUN-2001
602852690F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4993894 5',
mRNA sequence.
ACCESSION BI087886
VERSION BI087886
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 771)
NIH-MGC http://mgc.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: Incyte Genomics, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1M11015 row: d column: 23

High quality sequence stop: 762.

Location/Qualifiers

FEATURES

source

1..771

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4993894"

/cell_line="MGC36"

/lab_host="DH10B"

/clone_lib="NIH MGC 10"

/note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: oligo dt.

Average insert size 1.5 Kb. Library prepared by Life

Technologies."

ORIGIN

Query Match 44.7%; Score 593.4; DB 4; Length 771;
Best Local Similarity 89.2%; Pred. No. 2.2e-168;
Matches 688; Conservative 0; Mismatches 68; Indels 15; Gaps 4;
QY 107 GAGAAAGTTAAAGAAATTTGAGATCGAATATAATGTAGAGCAGAGACCTTCAAAACACT 166
DB 1 GAGAAAGTTAAAGAAATTTGAGATCGAATATAACGTTAGAGCAGAGAGCTTCGAAACACT 60
QY 167 GCACCTCTGGGCGCTCTCAGCCAAATGGATGCGCTTGGACTCTCCCTTTTAGGACCTCTA 226
DB 61 GGACCTCTGGGCGCTCTCAGCCAAATGGATGCGCTTGGACTCTCCCTTTTAGGACCTCTA 120
QY 227 GCAGCTATAATATTTTACTCTCTTTGACCTGTATCTTCAACTTCTCTGTTAGTTT 286
DB 121 GCAGCTATAATATTTGCTACTCTCTTTGGACCTGTATCTTTAACTCTCTTTTAACTTT 180
QY 287 GTCTCTTCAGAAATTTGAAGCTGTAAAGCTACAATAGTTCTTCAAAATGGAACCCAGATG 346
DB 181 GTCTCTTCAGAAATTTGAAGCTGTAAACTA-----CAATGAGCCCAAGATG 228
QY 347 CAGTCCATGACTAAATATTAACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGAT 406
DB 229 CAGTCCAGACTAAGATCTACCGCAGACCCCTGGACCGGCTGCTAGCCACGATCTGAT 288
QY 407 GTTAATGACATTGAAGTCAACCTCCGAGGAAATCTCACTGCACAAACCCCTACTACAC 466
DB 289 GTTAATGACATCAAGGCAACCCCTCTCTGAGGAAATCTCAGCTGCACAACTCTTACTACGC 348

[illegible]

RESULT 13	BC035153	BC035153	4204 bp	mRNA	linear	HTC 19-NOV-2000
LOCUS						
DEFINITION						
Homo sapiens hypothetical protein FLJ22313, mRNA (cdna clone IMAGE:5265109) . with apparent retained intron.						

ORGANISM

1 (bases 1 to 4204)

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 4204)

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L.H., Shenmen, C.M., Schuler, G.D., Altshul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stackleton, M., Soares, M.B., Bonaldo, M.F., Casavent, I.L., Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Saha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worthy, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Maman, A., Rodrigues, S., Sanchez, A., Whitting, M., Maman, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywicki, M.I., Skalska, U., Smalau, D.E., Schnerch, A., Schein, J.E., Jones, S.I., and Marra, M.A.

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D.,
cDNA Library Preparation: Michael J. Brown,
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Con-
DNA Sequencing by: Institute for Systems Bi-
<http://www.systemsbioology.org>
contact: amadan@systemsbioology.org
Anup Madan, Jessica Fahey, Erin Heiton, Mar-
Madan, Stephanie Rodrigues, Amy Sanchez, and

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 73 Row: k Column: 22
 This clone has the following problem: retained intron.

FEATURES	SOURCE
----------	--------

```

Location/Qualifiers
1. .4204
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5265109"
/tissue_type="Brain, hippocampus"
/clone_lib="NIH_MGC_95"
/lab_host="DH10B"
notes="Vector: pBluescript"

```

ORIGIN

Query Match	44.5%;	Score 592;	DB 3;	Length 4204;
Best Local Similarity	81.1%;	Pred. No. 9.9e+168;		
Matches 755;	Conservative	0;	Mismatches 128;	Indels 48; Gaps 4;
QY	400	CTCTGATGTTAAATGACATTGAAAGTCAACCCCTCCGAGGAAATCTCAACTGCGACAAACCCCT	459	
DB	887	CAC TGAATTAGTTGAGACACAGGCACCCCTACGGAGGAAATCTCAACTGCGACAAACCCCT	946	
QY	460	ACTPACACTCCAAATTCAGTTAGGAAGCAAGTTAGACAGTGTCTAGCCAAACCTTCCCAACAGT	519	
DB	947	ACTGTGCCCAATTCACGACGGAAGCAGTTAGAGCGGTGGTCA---AACCTCCCAATAGC	1003	
QY	520	ACTTTGGGTTTCTCGTTTGAGAGGGTGGACTGAGAGACAGGACTAGCTGGATTTCCTTAGGC	579	
DB	1004	ACTTTGGGTTTCTCGTTTGAGAGGGGGTACTGTAGAGACAGGACTAGTTTGGATTTCCTTAGGC	1063	
QY	580	TGACTAAGAATCCNAAAGCCTANCTGGGAAGGTGACCGCATCCATCTTTAAACATGGGGC	639	
DB	1064	TGGCTAAGAATCCCTTAAGCCTAGCTGGGAAATTGACCACTCCACCTTTAAACACGGGC	1123	
QY	640	TTTGCAACTTAGCTTCACCCGGACCAATC-----AGAGAGCTCACTAAATGCTAAT	690	
DB	1124	TTTGCAATTTAGCTTCACCCGCCCAATCAGGTAGTAAAGAGAGCTCACTAAATGCTAAT	1183	
QY	691	CAGGCAAAACAGGAGGTAAAGCAATAGCCAATCATCTATTGCCTCAGAGCACAGCGGA	750	
DB	1184	TAGGGAHAACAGGAGGTAAAGATAGCCATCATCTATCGCTCGAGAGCACACAGGA	1243	
QY	751	AGGACAAAGATTGGGATATAAATCTAGGCATTCAAGCCAGCAACAGCAACCCCTTTGGG	810	
DB	1244	GGGACAATGATCAGGATATAAACCCAGGCATTCAAGCCAGCGGTAGCTACCCCTTTTGGG	1303	
QY	811	TCCCTTCCCATTTGATGGGAGCTCTGTTTTCACCTCTATTTCACCTCTATTAAATCATGCAA	870	
DB	1304	TCCCTTCCCTTTGATATGGAAGCTCTGTTTTCACCTCTATTAAATCTTGCAATTGCA-----	1358	
QY	871	CTGCACTCTTCTGCTCCGTGTTTTTATGGCTCAAGCTGAGCTTTTGTTCGGCATCCACC	930	
DB	1359	---CACTTTTCTGGTACGTTGTGTGCACAGCTCAAGCTGAGCTTTTCGTTCACGTCACCC	1415	
QY	931	ACTGCTGTTTGGCAACCGTCAAGACCCGCTGTGACTTCCATCCCTTTGGATCCAGGAGA	990	
DB	1416	ACTGCTGCTGCGCGTGTACAGACCCACAGCTGACTTCCATCCCTCTG-----	1464	

```
QY 991 GTGTCCACTGTCTCTGATCCAGCGAGGTACCCATTGGCCACTCCCGATCAGGCTAAAGG 1050
Db 1465 -----GATCCAGCGAGGCGCCCATTTGCTGCTCCATCGGCTAAAGG 1507
QY 1051 CTTGCCATTGTTCTGTCATGGCTAAAGTGCCTGGGTTTGTCTTAATAGAACTGAACACTGG 1110
Db 1508 CTTGCCCTTGTCTCTGAGGGCTAAGTGCCAGGTTCTGCTTAATCTAGCTAAACACTAG 1567
QY 1111 TCACTGGGTTCCATGGTTCTTCCATGACCCACCGGCTTCTTAATAGAGCTATAAACACTCA 1170
Db 1568 TCACTGGGTTCCAGGATTTCTTCCATGACCCACAGCTTCTTAATAGAGCTATAAACACTCG 1627
QY 1171 CCGCATGGCCCAAGATTCCTTGGTATCTGTGAGGCCAAGAACCCCGAGCTCAGAGA 1230
Db 1628 CCACTTGGGCCAAGATTCCTTGGTATCTGTGAGGCCAAGAACCCCGAGCTCAGAGA 1687
QY 1231 ANGTGAGGCTTGGCACCATTGGGAAGTGGCCCACTGCCATTGTTGGTAGGGGCCCAAC 1290
Db 1688 ACACAGACTTGGCACCATTCTGGAGTGGCCCGTCAACCATCTTGGAGGGACCTGCCAC 1747
QY 1291 CATCTTGGGAGCTGTGGGCAAGGATCCCC 1321
Db 1748 CATCTTGGGAACCTTGGGCAAGGACCCCC 1778

RESULT 14
LOCUS CN272394 719 bp mRNA linear EST 16-MAY-2004
DEFINITION 17000600005410 GRN_PRENU Homo sapiens cDNA 5', mRNA sequence.
ACCESSION CN272394
VERSION CN272394.1 GI:47288808
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 719)
Li, Y., Xu, C., Fang, R., Zhang, S., Lei, S., Murage, J., Fisk, G. J.,
Lebkowski, J. and Stanton, L. W.
Transcriptome characterization elucidates signaling networks that
control human ES cell growth and differentiation
Nat. Biotechnol. 22 (6), 707-716 (2004)
CONTACT: Brandenberger R
Regenerative Medicine
Genon Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbrandenberger@genon.com
Insert Length: 719 Std Error: 0.00.
FEATURES
source
1..719
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cell, retinoic acid and
mitogen-treated hES cell line H7"
/clone_lib="GRN_PRENU"
/notes="oligo dT primed, full-length enriched cDNA library
from hES cell line H7 (p29) maintained in feeder-free
conditions. Embryoid bodies were generated in the presence
of all-trans retinoic acid and mitogens."

Query Match 42.2%; Score 561; DB 7; Length 719;
Best Local Similarity 85.4%; Pred. No. 1.5e-158;
Matches 657; Conservative 0; Mismatches 62; Indels 50; Gaps 1;

QY 95 GGAATCATTTACTGAGAAGTTAAAGAAATTTGAGATCGGAATAAATCTAGAGCAGAGGAC 154
Db 1 GGAATCGTCACTGAGAAGTTAAAGAAATTTGAGATCGGAATAAATCTAGAGCAGAGGAG 60
```

```
QY 155 CTTCAAAACACATGCACCCCTGGGSCCTCCTCAGCCAATGATGCCCTGGACTCTCCCTTC 214
Db 61 CTTGGAACACATGAGCCCTGGGSCCTCCTCAGCCAATGATGCCCTGGACTCTCCCTTC 120
QY 215 TTAGGACCTCTAGCAGCTATAATATTTTTACTCCTCTTTGGACCCCTGTATCTTCAACTTC 274
Db 121 TTAGGACCTCTAGCAGCTATAATATTTGCTACTCCTCTTTGGACCCCTGTATCTTAACTTC 180
QY 275 CTTGTTAAAGTTTGTCTCTTCCAGAAATTGAAGCTGTAAAGCTACAAATAGTTCTTCAATG 334
Db 181 CTTGTT-----AATG 190
QY 335 GAACCCACAGATCAGTCCATGACTAAATCTACCGTGGACCCCTGGACCGGCTGTCTAGA 394
Db 191 GAGCCCAAGATGAGTCCCAAGACTAAGATCTACCGCAGACCCCTGGACCGGCTGTCTAGC 250
QY 395 CTATGCTCTGATGTTTAATGACATTTGAAGTCAACCCCTCCGAGGAAATCTCAATGCACAA 454
Db 251 CCAGATCTGATGTTTAATGACATCAAAAGGCACCCCTCCTGAGGAAATCTCAGCTGCACAA 310
QY 455 CCCCTACTACATCCAAATTCAGTAGGAAGCAGTTAGAGCAGTTGTGAGCCCAACCTCCCCA 514
Db 311 CCTCTACTACGCCCAATTCAGCAGGAAGCAGTTAGAGCGGTCTGCGGCCAACCTCCCCA 370
QY 515 ACAGTACTTTGGGTTTCTCTGTTGAGAGGCTGGACTGAGAGACAGGACTAGCTGGATTTC 574
Db 371 ACAGCACTTAGGTTTCTCTGTTGAGATGGGGAAGTCTGAGAGACAGGACTAGCTGGATTTC 430
QY 575 TAGGCTGACTAAGAAATCCCAAGCCCTTANCTGGGAAGTGCACCGCATCTCTTTAAACAT 634
Db 431 TAGGCTGACTAAGATCCCTTAAAGCTAGCTGGGAAGTGCACACATCCACCTTTAAACAC 490
QY 635 GGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGCTCACTAAAATGCTTAATCAGG 694
Db 491 GGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGCTCACTAAAATGCTTAATAGG 550
QY 695 CAAAACAGAGGTAAAGCAATAGCAATCATCTATTGCTGAGACAGCAGCGGGAAGGA 754
Db 551 CAAAGACAGGAGGTAAAGAAATAGCCAAATCATCTATTGCTGAGACAGCAGCGGGAAGGA 610
QY 755 CAAGGATTCGGGATATAAACTCAGGCAATCAAGCCAGCAACAGCAACCCCTTTGGGTCCC 814
Db 611 CAATGATCGGATATAAACCCAGTCTTCGAGCCGCAACGCAACCCCTTTGGGTCCC 670
QY 815 CTCCCATTTGATGGGAGCTCTGTTTCACTCTATTTCCTCTATTATAAT 863
Db 671 CTCCCTTTGATGGGAGCTCTGTTTCACTCTATTTCCTCTATTATAAT 719

RESULT 15
LOCUS BX459153/c 900 bp mRNA linear EST 05-MAY-2004
DEFINITION BX459153 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0E013Y120
3-PRIME mRNA sequence.
ACCESSION BX459153
VERSION BX459153.2 GI:47051796
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 900)
Li, W. B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
COMMENT On May 22, 2003 this sequence version replaced gi:31021096.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
```

was not normalized. Library was constructed by Life Technologies, a division of Invitrogen.

This sequence belongs to sequence cluster 4215.1

For more information about this cluster, see

<http://www.genoscope.cns.fr/cdna?s=CS0DE013BE1ONP1&c=4215.1>.

FEATURES

source

1..900
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE013I120"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 40.9%; Score 544; DB 5; Length 900;
Best Local Similarity 77.5%; Pred. No. 2.4e-153;
Matches 631; Conservative 49; Mismatches 120; Indels 14; Gaps 3;

Qy 1 TCAAAATCGAAGAGCTTTAGACTGCTTAACCGCAAAAGAGGGGAACTGTTTATTTT 60
Db TCAAAATCGAAGAGCTTTAGACTGCTTAACCGCTGAAGAGGGGAACTGTTTATTTT 743

Qy 61 AGGGAGAGATGCTTTAGTATGTTAATCAATCTGGAATCATTTACTGAGAAAGTTAAGA 120
Db AGGGAGAGATGCTTTAGTATGTTAATCAATCTGGAATCATTTACTGAGAAAGTTAAGA 693

Qy 121 AATTGAGATCGAATAATAATGTAGACGAGAGGACCTTCAAAACACTGCACCTGGGCGCT 180
Db AATTGAGATCGAATAATAAAGTAGACGAGAGGAGCTTCGAAACACTGGACCTGGGCGCT 623

Qy 181 CCTCAGCAATGGATGCCCTGGACTCTCCCTCTTTAGGACCTCTAGCAGCTATAATTT 240
Db CCTCAGCAATGGATGCCCTGGACTCTCCCTCTTTAGGACCTCTAGCAGCTATAATTT 563

Qy 241 TTTACTCTCTTTGGACCTGTATCTTCAACTTCTCTTTAGTTTGTCTCTTCAGAAAT 300
Db TTTACTCTCTTTGGACCTGTATCTTCAACTTCTCTTTAGTTTGTCTCTTCAGAAAT 505

Qy 301 TGAAGCTGTAAAGCTACAAATAGTCTTCAATAGGAACCCAGATGCAGTCCACTGACTAA 360
Db TGAAGCTGTAAAGCTACAAATAGTCTTCAATAGGAACCCAGATGCAGTCCACTGACTAA 457

Qy 361 AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGTATGTTAATGACATGA 420
Db AATCTACCGTGGACCCCTGGACCGGCTGCTAGACTATGCTCTGTATGTTAATGACATGA 497

Qy 456 SACTACCGCAGACCCCTGGACCGGCTGCTAGACTATGCTCTGTATGTTAATGACATGA 497
Db SACTACCGCAGACCCCTGGACCGGCTGCTAGACTATGCTCTGTATGTTAATGACATGA 497

Qy 421 AGTCACCCCTCCGAGGAAATCTCACTGACACACCCCTACTACTACTCCAAATTCAGTAGG 480
Db AGTCACCCCTCCGAGGAAATCTCACTGACACACCCCTACTACTACTCCAAATTCAGTAGG 480

Qy 396 ASGCACCCCTCCGAGGAAATCTCACTGACACACCCCTACTACTACTCCAAATTCAGTAGG 480
Db ASGCACCCCTCCGAGGAAATCTCACTGACACACCCCTACTACTACTCCAAATTCAGTAGG 480

Qy 481 AAGCAGTTAGAGCAGTGTGTAGCAGCAACCTCCCAACAGTACTGGGTTTCTCTGTGTAGA 540
Db AAGCAGTTAGAGCAGTGTGTAGCAGCAACCTCCCAACAGTACTGGGTTTCTCTGTGTAGA 540

Qy 336 AAGCAGTTAGAGCAGTGTGTAGCAGCAACCTCCCAACAGTACTGGGTTTCTCTGTGTAGA 540
Db AAGCAGTTAGAGCAGTGTGTAGCAGCAACCTCCCAACAGTACTGGGTTTCTCTGTGTAGA 540

Qy 541 GGGTGGACTGAGACAGGACTAGTGGATTTCTAGGCTGACTAAGAAATCCCAAGGCTT 600
Db GGGTGGACTGAGACAGGACTAGTGGATTTCTAGGCTGACTAAGAAATCCCAAGGCTT 600

Qy 276 TGGGSCACTSAGASACAGSACTAGCTSSATCMGCCMCCGCCCTCCCAAGGCTT 617
Db TGGGSCACTSAGASACAGSACTAGCTSSATCMGCCMCCGCCCTCCCAAGGCTT 617

Qy 601 ANCTGGGAGGTGACCGCTATCTTTAAACATGGGCTTGAATTTCTAGGCTGACTAAGAAATCCCAAGGCTT 600
Db ANCTGGGAGGTGACCGCTATCTTTAAACATGGGCTTGAATTTCTAGGCTGACTAAGAAATCCCAAGGCTT 600

Qy 216 MGCTGGAGGTGACCGCTATCTTTAAACATGGGCTTGAATTTCTAGGCTGACTAAGAAATCCCAAGGCTT 600
Db MGCTGGAGGTGACCGCTATCTTTAAACATGGGCTTGAATTTCTAGGCTGACTAAGAAATCCCAAGGCTT 600

Qy 661 ACCAATCAGAGCTCAATAAATGCTTAATCAGGCAAAACAGGAGGTAAAGCAATGAGCC 720
Db ACCAATCAGAGCTCAATAAATGCTTAATCAGGCAAAACAGGAGGTAAAGCAATGAGCC 720

Qy 156 SCCAATCAGAGCTCAATAAATGCTTAATCAGGCAAAACAGGAGGTAAAGCAATGAGCC 720
Db SCCAATCAGAGCTCAATAAATGCTTAATCAGGCAAAACAGGAGGTAAAGCAATGAGCC 720

Qy 721 AATCATCTATTGCTGAGACACAGCGGGAAGGACAGGATTTGGGATATAAATCTCAGGCA 780
Db AATCATCTATTGCTGAGACACAGCGGGAAGGACAGGATTTGGGATATAAATCTCAGGCA 780

Db 96 MVCCATSCACTGCTGAGAGCACRGACGAGGAGACVATGCGGGAGRTAAACCAAGTC 37
Qy 781 TTCAAGCCAGACACAGCAACCCCTTTGGTCCC 814
Db 36 BCCGCGCCGGCTACGGCKCCCCCTXTGGTCCCC 3

RESULT 16

BX953822/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

MIPS

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

Sequenced by EMBL (European Molecular Biology Laboratories,

Heidelberg/Germany) within the cDNA sequencing consortium of the

German Genome Project.

No sl sequence available.

This clone (DKFZp781M064) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Location/Qualifiers

1..1002

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFZp781M064"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="781 (synonym: hlcc4)"

/note="Vector: pSPORT1_Sfi; Site_1: SfiIA; Site_2: SfiIB;

cDNA-collection"

ORIGIN

Query Match 37.7%; Score 500.6; DB 5; Length 1002;

Best Local Similarity 87.6%; Pred. No. 3.8e-140;

Matches 641; Conservative 0; Mismatches 72; Indels 19; Gaps 8;

Qy 549 TGAGAGACAGGACTAGCTGGATTCTTAGGCTGACTAAGAAATCCCAAGGCTTCTGGA 608

Db 922 TGAGAGACAGGACTAGCTGGATTCTTAGGCTGACTAAGAAATCCCAAGGCTTCTGGA 863

Qy 609 AGGTGACCGCTATCTTTAAACATGGGCTTGAATTTCTAGCTCACCACCGCAATCA 668

Db 862 A-GTACCGCTATCTTTAAACATGGGCTTGAATTTCTAGCTCACCACCGCAATCA 808

Qy 669 GAGAGCTCACTAAATGCTTAATCAGGCAAAACAGGAGTAAAGCAATGCAATCATCT 728

Db 807 -----CTCACTAAATGCTTAATCAGGCAAAACAGGAGTAAAGCAATGCAATCATCT 753

Qy 729 ATTGCTCAGACGACGCGGAGGACAGGATTTGGGATATAAATCTCAGGATTTCAAGCC 788

Db 752 ATTGCTCAGACGACGCGGAGGACAGGATTTGGGATATAAATCTCAGGATTTCAAGCC 694

Qy 789 AGCAACAGCAACCCCTTTGGTCCCTCCCATTTGATGGGAGCTCTCTTTTCACTCAT 848

Db 848

```
Db 693 AGCAATGGCAACCCCTTTGGGTCCTCCCTCGTATGGAGCTGTGTTTCACCTAT 634
Qy 849 TTCACTCTATTAATCATGCACTGCACTCTCTCGTCCGTGTTTATGGCTCAAGCT 908
Db 633 TTCACACTATTAATCTTGAACCTGCACCTCTCTCGTTCGTGTTGTTACTGCTGAGCT 574
Qy 909 GAGCTTTTGTTCGCATCCACCACTGCTGTT-TGCCACGTCACACACCGCTGCTGACT 967
Db 573 GAGCTTTGCTCACTGTCACCACTGCTGTTTGGCGCGGTAGCAGACTGCTGCTGACT 514
Qy 968 TCCATCCCTTTGGATCCAGCAGAGTGCACCTGCTGCTCTGATCCAGCGAGGTA-CCCAT 1026
Db 513 TACATTTCTCCGATCCAGCAGAGGTGTCACCTGCTGCTCTGATCCAGCGAGGACCCCTT 454
Qy 1027 TGCACCTCCCGATCAGCTAAAGGCTTGCCA-TTGTTCCTGCAATGCTAAGTCCCTGGGT 1085
Db 453 TCGCGCTCCAGATCGGCTAAAGGCTTGCCATTTGTTCTGTCACGCTAAGTCCCTGGGT 394
Qy 1086 TTGCTCTAATAGAACTGAACACTGGTCACTGGGTTTCATGGTCTCTTCCATGACCCACG 1145
Db 393 TTGCTCTAATAGAACTGAACACTAGTCACTGGGTTCCATGGTGTCTTCGCTGACCCACG 334
Qy 1146 GCTTCTAATAGAACTGAACACTCAGCAGTCCAGGCTGCGCAAGATTCCTTGGTATCTGT 1205
Db 333 ACTTCTAATAGAACTGAACACT------GCATGGCCCAAGATTCCTTGAATCCGT 278
Qy 1206 GAGGCCAAGAACCCAGGTCAGAGAANGTGAGGCTTGCCACCAT-TTGGGAAGTGCCCA 1264
Db 277 GAGGCCAAGAACCCAGGTCAGAGAACAGAGGCTTGCCATCATCTTTGGAAGTGCCCA 218
Qy 1265 CTGCCATTTGG 1276
Db 217 CGACCATTTG 206
```

```
RESULT 17
AG113694/c 679 bp DNA linear GSS 03-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-120G11.R, genomic survey sequence.
DEFINITION AG113694
ACCESSION AG113694
VERSION AG113694.1 GI:16734213
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE 1
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
TITLE Totoki, Y., Watanabe, H. and Sakaki, Y.
JOURNAL BAC end sequences of Library PTB
AUTHORS Unpublished
TITLE 2 (bases 1 to 679)
JOURNAL Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
AUTHORS Totoki, Y., Watanabe, H. and Sakaki, Y.
COMMENT Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
E-mail: chimpes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
Location/Qualifiers
1. .679
/organism="Pan troglodytes"
```

```
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-120G11.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
```

ORIGIN

```
Query Match 36.2%; Score 480.8; DB 9; Length 679;
Best Local Similarity 86.7%; Pred. No. 3.5e-134;
Matches 552; Conservative 0; Mismatches 83; Indels 2; Gaps 2;
Qy 673 GCTCACTAAATGCTAAATCAGGCAAAACAGAGAGTAAAGCAATAGCCAAATCATCTATTG 732
Db 678 GCTCGGTAGATGCTAAATAGGCAAAACAGAGAGTAAAGCAATAGCCAAATCATCTATTG 619
Qy 733 CTTGAGAGCACAGCGGAAAGGACAGGATTGGGATATAAACTCAGGCATTCAAGCCAGCA 792
Db 618 CGTGAGAGCACAGTGTAAAGAAAGGGTCGGATATTAGCCCGGCATTTCGAGCCGGCA 559
Qy 793 AC-AGCAACCCCTTTGGGTCCCTCCCATTTGTATGGAGCTCT-GTTTTCACTCTATT 850
Db 558 ACGGGCAGCCCTTTGGGTCCCTCCCTTTGTATGGCAGCTCTGCTTTTCACTCTATT 499
Qy 851 CACTCTATTAAATCATGCACTGCACTCTTCTGGTCCGTGTTTTTATGCTCAAGCTGA 910
Db 498 CACTATTATTAGATCTTGCAGCTGCACTCTTCTGGTCCGTGTTTTTATGCTCAAGCTGA 439
Qy 911 GCTTTTGTTCGGCATCCACCACTGCTGTTTGGCACCCTCACAGACCCGCTGCTGACTTCC 970
Db 438 GCTTTTGTTCGGTCCGCTCCGCTGCTGTTTACCGCGCTCCAGACCCGCTCACTGACTTCC 379
Qy 971 ATCCCTTTGGATCCAGCAGAGTGTCCATGTGCTCTCTGATCCAGCAGGTACCCATTGCC 1030
Db 378 ATCCCTCTGGATCCGCGAGGCTGCGCTGTGCTACTGATCCAGCAGGACCCCAATTGCT 319
Qy 1031 ACTCCCGATCAGGCTTAAAGGCTTGCCATTTGCTTCGATGGCTTAAGTCTGGGTTGTC 1090
Db 318 GCTCCCGATCGGGCTTAAAGGCTTGCCATTTGCTTCGATGGCTTAAAGTCTGGGTTGTC 259
Qy 1091 CTAATAGAACTCAACACTGCTCAGTGGTTCATGGTCTCTTCCATGACCCACGCTTC 1150
Db 258 GTAATGAGCTCAACACTAGTGTGCTGAGTTCACAGTTCTTCCGTGACCCACGCTTC 199
Qy 1151 TAATAGAGTATAACACTCACCGCATGCGCCCAAGATTCCATTCTTGGTATCTGTGAGGC 1210
Db 198 TAATAGAACTGTAAACGCTCACCGCATGCGCCCAAGATTCCGTTCCTTGGAAATCCGTGAGGC 139
Qy 1211 CAAGAACCCCGAGTCAAGAAAGTGGAGCTTGCCACCATTTGGGAAGTGGCCCACTGCCA 1270
Db 138 CAAGAACCCCGAGTCAAGAAACACGAGGCTTGCCACCATTTGGGAAGTGGCCCGCTGCCA 79
Qy 1271 TTTTGTAGTCGCGCCACCACTTCTTGGGAGCTGTGG 1307
Db 78 TTTTGAAGTGGCCCGCCCACTTCTTGGGAGCTCTCG 42
```

RESULT 18

BM918330/c

LOCUS

DEFINITION

5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

```
BM918330 999 bp mRNA linear EST 12-MAR-2002
AGENCOURT_6708649 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5747240
5', mRNA sequence.
BM918330
BM918330.1 GI:19368709
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 999)
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
```

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLAM12773 row: b column: 09
 High quality sequence stop: 685.
 Location/Qualifiers
 1. 999
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5747240"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_120"
 /note="Organ: pooled pancreas and spleen; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of spleen and pancreas from 28 yo male. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-2.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 025. Note: this is a NIH_MGC Library."
 ORIGIN

Query Match 36.2%; Score 480.8; DB 4; Length 999;
 Best Local Similarity 83.1%; Pred. No. 4e-134;
 Matches 621; Conservative 0; Mismatches 110; Indels 16; Gaps 6;
 QY 549 TGAGAGACGAGCTAGCT-GGATTTCTAGCTGACTAAGATCCCAAGCCTTANTGGG 607
 Db 886 TGAGAGACGAGCTAGCTGGATTTCTAAGCGGAGCTAAGATCCCTTAGCCCTTAGTAGG 827
 QY 608 AAGGTGACCCGATCCATCTTTAA--CATGGGGCTTGCAACTAGCTCACACCCGACCAAT 666
 Db 826 AAGGTGACCCGATCCATCTTTAA--CATGGGGCTTGCAACTAGCTCACACCCGACCAAT 767
 QY 667 C-----AGAGAGCTCACTAAATGCTTAATCAGGC-AAAAACAGGAGGTTAAGCAAT 716
 Db 766 CAGGTAGTAAAGAGAGCTCACTAAATGCCAATCAGACAAAAACAGGAGTTAAGCAAT 707
 QY 717 AGCAATCATCTATTGCTCAGAGCAGCGGAGCAGCAAGGATTTGGGATATAAATCA 776
 Db 706 AACCAATCATCTATTGCTCAGAGCAGCGGAGGAGCAATGATCAGGATATAAGCCA 647
 QY 777 GGCAATCAAGCCAGCAACACCCCTTTGGGTCCCTCCCATTTGTTATGGAGCTCTG 836
 Db 646 GGCATTCAGCGGCAACGCTACCTCTTTGGCTCCCTCCCTCTTTGTTATGGAGCTGTG 587
 QY 837 TTTTCACTATTCTACTATTAAATCATGCAACT---GCACTCTTCTGGTCCGTGTTT 893
 Db 586 TTTTCACTATTAAATCTTTGTAATGTCATCTCTTCTGTCGCTGTTTGTACCGTGT 527
 QY 894 TTTATGCTCAAGCTAGCTTTGTTGCGCATCCACACTGCTGTTTGCCACCGTCACAG 953
 Db 526 GTTACCGCTCAAGTAAAGCTTTTCGCTGTTGTCACACACTGCTGTTGCGGAGTCGAG 467
 QY 954 ACCCGTGTGCTTCCATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCTGATCA 1013
 Db 466 ACCCGCGCTGACTTCCACCCCTCCGGATCTGCGAGGGTGTCTGCTGTCTCTGATCA 407
 QY 1014 GCGAGTACCATTGCCATCTCCGATCAGGCTAAGGCTTGCATTTCTCTGCAATGGCT 1073
 Db 406 GCGAGGCGCCCACTGTGCTCCGATCCGGCTGAAGGCTTGCCTGTTGTTCTCTGACGCG- 348
 QY 1074 AAGTGCTGGTGTGCTTAATAGAACTGAACACTGCTGCTACCTGGTTCATGTTCTCTT 1133
 Db 347 AAGTGCTGGTGTGCTTAATGAGATGAACACTAGTCTGCTGGGTTCTACGGTTCTCTT 288

QY 1134 CCATGACCCACGGCTTCTAATAGAGCTATAACAATCACCAGCATGGCCCAAGATTCCATTC 1193
 Db 287 CCATGACCCACGGCTTCTAATAGAGCTATAACAATCACCAGCATGGCCCAAGATTCCATTC 228
 QY 1194 CTTGGTATCTGTAGGCCAAGAACCCAGGTGAGAGAANGTGAAGTTCGCCACATTGG 1253
 Db 227 CTTGGAATCCGTGGGCCAAGAACCCAGGTGAGAGAANGTGAAGTTCGCCACATTGG 168
 QY 1254 GAAGTGGCCCACTGCCATTTTGGTAGC 1280
 Db 167 GAAGCAACCCGCCACCATGTTGGAGC 141
 RESULT 19
 AG134524
 LOCUS 683 bp DNA linear GSS 04-NOV-2001
 DEFINITION Pan troglodytes DNA, clone: PTB-147110.R, genomic survey sequence.
 ACCESSION AG134524
 VERSION AG134524.1 GI:16664202
 KEYWORDS GSS.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 REFERENCE 1
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE BAC end sequences of Library PTB
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 683)
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpansegsc.riken.go.jp, URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the RED process and may have higher chance of clone tracking errors.
 PRIMERS
 Sequencing: M13Rev
 LIBRARY Vector : pKS145
 R.Site 1 : SacI
 R.Site 2 : SacI.
 FEATURES
 source Location/Qualifiers
 1. 683
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /clone="PTB-147110.R"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_lib="PTB Chimpanzee Male BAC Library"
 ORIGIN
 Query Match 35.8%; Score 475.2; DB 9; Length 683;
 Best Local Similarity 89.6%; Pred. No. 1.8e-132;
 Matches 554; Conservative 0; Mismatches 60; Indels 4; Gaps 4;
 QY 666 TCAGAGAGCTCACTAAATGCTTAATCAGCAAAAACAGGAGTAAAGCAATAGCCAATCA 725
 Db 58 TCNAGAGCTCACTAAATGCTTAATAGCAAAAACAGGAGTAAAGCAATAGCCAATCA 117
 QY 726 TCTATTGCTGAGACACAGCGGAGGACAGGATTTGGGATATAAACTCAGGCAATTC 785
 Db 118 TCTATTGCTGAGACACAGTGGGAGGACAGGATCGGATATAAACCCAGGCAATTC 177
 QY 786 GGCACCAACAGCAACCCCTTTGGTCCCTCCCATTTGATGGAGCTCTGTTTCACTC 845

```
Db 178 GCAGCAACGGCAACCCCTTTGGGT-CCCTCCCTTGATGGAGCTCTGTTTCACTC 236
Qy 846 TATTCACTCTATAATATGCAACTGCACTCTTCGTGTCCTGTTTATGGCTCAA 905
Db 237 TATTCACTCTATAATATGCAACTGCACTCTTCGTGTCCTGTTTATGGCTCAA 296
Qy 906 GCTGAGCTTTGTTCGCCATCCACCACTGCTG-TTTGCCACCGTCACAGACCGCTGCTG 964
Db 297 GCTGAGCTTTCTTCGCCATCCACCACTGCTGTTTTCGCCCGCTCGAGACCGCGCTG 356
Qy 965 ACTTCCATCCCTTTGGATCCAGCAGAGTGTCCTGCTGCTGATCCAGCGAGGTACCC 1024
Db 357 ACTTCCATCCCTCCAGATCCAGCAGAGTGTCCTGCTGCTGATCCAGCGAGCGCC 416
Qy 1025 ATTGCCACTCCGATCAGGCTAAAGCTTGCCATTTGCTGATGCTGCTGCTGCTGCTG 1084
Db 417 ATTGCCGCTCCTGATCGGCTAAAGCTTGCCATTTGCTGATGCTGCTGCTGCTGCTG 476
Qy 1085 TTTGTCTTAATAGACTGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1144
Db 477 TTTGTCTTAATAGACTGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 536
Qy 1145 GGCTTCTTAATAGACTGATACACTCACCGCATGGCCCAAGATTCCATTCT-TGCTATCT 1203
Db 537 GACTTCTTAATAGACTGATACACTCACCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 596
Qy 1204 GTGAGGCCAA-GAACCCCGAGTGCAGAGAAAGTGAGGCTTGCCACCAATTGGGAAAGTGCC 1262
Db 597 ATGAGGCCAATGAACCCCGAGTGCAGAGAAACACGAGGCTTGCCACCACTTGGAAAGTGCC 656
Qy 1263 CACTGCCATTTGGTAGC 1280
Db 657 CACCACCATCTGGGGAAC 674

RESULT 20
AG049953
LOCUS Pan troglodytes DNA, clone: PTB-030D19.F, linear GSS 02-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-030D19.F, genomic survey sequence.
ACCESSION AG049953
VERSION AG049953.1 GI:16586845
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library PTB
Unpublished
2 (bases 1 to 669)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1. .669
/organism="Pan troglodytes"
/mol_type="genomic DNA"

FEATURES
source
```

```
/db_xref="taxon:9598"
/clone="PTB-030D19.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
```

ORIGIN

```
Query Match 35.7%; Score 474; DB 9; Length 669;
Best Local Similarity 86.8%; Pred. No. 4.1e-132;
Matches 548; Conservative 0; Mismatches 71; Indels 12; Gaps 2;

Qy 669 GAGAGCTCACTAAATGCTAAATCAGGCAAAACAGGAGGTAAAGCAATAGCCAAATCATCT 728
Db 49 GAGAGCTCACTAAATGCTAAATGCTAAATGCTAAATGCTAAATGCTAAATGCTAAATCATCT 108
Qy 729 ATTGCTGAGAGCAGCGGGAAGGACAAAGGATTTGGATATATAACTCAGGCATTTCAAGCC 788
Db 109 ATTGCTGAGAGCAGCAGGAGGAGCAATGATCGGATATAAAACCCAGCATTCGAGCT 168
Qy 789 AGCAACAGCAACCCCTTTGGGTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 848
Db 169 GGCACCGGCTACCTTTCTTTGGGTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 218
Qy 849 TTCACCTCTATTAATCATCACTGACATC--TTCTGTGTCCTGTTTATGGCTCAAG 906
Db 219 TTCACCTCTATTAATCTTGCAACTGCACCTCTTTTATGTTCCATGTTTGTACGCTCGAG 278
Qy 907 CTGAGCTTTGTTTGGCCATCCACCACTGCTGTTTGGCCCGTCACAGACCCCGCTGCTGAC 966
Db 279 GTGAGCTTTGATGCGGTCCCGTCCAGTGTCTGTTTGTGCGTCGACAGCCACCGCTGAC 338
Qy 967 TTCATTCCTTTGGATCCAGAGAGTCCACTGTGCTCTCTGATCTCTGATCAGGAGGTACCCAT 1026
Db 339 TTCATTCCTCTGGATCTGCGGGTGTCCGTGTCTCTGATCAGGAGGTGCCCCAT 398
Qy 1027 TGCCACTCCCGATCAGGCTAAAGGCTTGCCATTTGTTCTGTCAGTGGCTAAGTGCCTGGTT 1086
Db 399 TGTGCTCTCTGATCGGATAAAGGCTTGCCATTTGTTCTGCAAGGCTAAGTGCCTGGTT 458
Qy 1087 TGTCTTAATAGAACTGAACACTGGTCACTGGGTTCATGGTTCCTCTTCCATGACCCACGG 1146
Db 459 CGTCTAATAGAGTGAACTAGTCACTAGTCACTGGGTTCATGGTTCCTCTTCCGTAATGG 518
Qy 1147 CTTCTAATAGAGTATAACACTCAGCGATGGCCCAAGATTCATTCCTCTTGGTATCTGTG 1206
Db 519 CTTCTAATAGAGTATAACACTCAGCGATGGCCCAAGATTCATTCCTCTTGGAAATCCGG 578
Qy 1207 AGCCCAAGAACCCCGAGTGCAGAGAGTGCAGGCTTGCCACCATTTTGGGAAGTGGCCACT 1266
Db 579 AGCCCAAGAACCCCGAGTGCAGAGAGTGCAGGCTTGCCACCATTTTGGGAAGTGGACACC 638
Qy 1267 GCATTTTGTAGCGGCCCCACCACTATCTTG 1297
Db 639 ACCATTTTGAAGACCGCCCGCCACCATGTTG 669

RESULT 21
AG116971
LOCUS Pan troglodytes DNA, clone: PTB-124K18.F, linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-124K18.F, genomic survey sequence.
ACCESSION AG116971
VERSION AG116971.1 GI:16737490
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library PTB
Unpublished
2 (bases 1 to 681)
```

AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
TITLE Totoki,Y., Watanabe,H. and Sakaki,Y.
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-chou,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:chimbpe@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
 Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
 was generated during the R&D process and may have higher chance of
 clone tracking errors.
PRIMERS
 Sequencing: -21M13
LIBRARY
 Vector : pKS145
 R.Site 1 : SacI
 R.Site 2 : SacI
FEATURES Location/Qualifiers
 source 1..681
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /clone="PTB-124K18.F"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
 Query Match 35.6%; Score 473; DB 9; Length 681;
 Best Local Similarity 88.8%; Pred. No. 8.2e-132;
 Matches 538; Conservative 0; Mismatches 56; Indels 12; Gaps 2;
 QY 671 GAGCTCACTAAATGCTAATCAGGCAAAACAGGAGTAAGCAATAGCCAAATCATCTAT 730
 Db 86 GAGCTCACTAAAGCTAATTAGGCAAAACAGGAGTAAGCAATAGCCAAATCATCTAT 145
 QY 731 TGCTGAGACACAGCGGGAGGACAGGATTTGGGATATAAACTCAGGCATTCAAGCCAG 790
 Db 146 TGCTGAGACACAGCAGGAGGAGCAATGATTGGGATATAAACCCAGGCATTCGAGCCAG 205
 QY 791 CAACAGCAACCCCTTTGGTCCCTCCCATTTGTATGGGAGCTCTGTTTTCACTATT 850
 Db 206 CAACGGCTACCCCTTTGGTCCCTCCCTTTGTATGGGAGCTCTGT-----TTT 255
 QY 851 CACTCTATTAAATCATGCAACTGCA--CTCTCTGGTCCGTGTTTTTATGGCTCAAGCT 908
 Db 256 CACTCTATTAAATCTTGCAACTGCACTCTCTCTGGTCCGTGTTGTATGGCTCAAGCT 915
 QY 909 GAGCTTTTGTTCGCCATCCACCACTGTGTTTGGCCACCGTCAAGCCCGCTCTGACATT 968
 Db 316 GAGCTTTTGGCTTGCCATCCACCACTGCTGTTTGGCGCGCGGAGACCGCTCTGACATT 375
 QY 969 CCAATCCCTTTGGATTCAGCAGAGTGTCCTCACTGCTCTCTGATCCAGCAGGTACCAATTG 1028
 Db 376 CCAATCCCTCCAGATCTGGCAGGGTGTCCTCACTGCTCTCTGATCCAGCAGGGGCCATTG 435
 QY 1029 CCACTCCGATCAGGCTAAAGGCTTCCATTGTTCTGATCGCTAAGTCCCTGGGTTTG 1088
 Db 436 CCCTCCCGATAGGGCTAAAGGCTTGCATTGTTCTGATGGCTAACTGCCAGGTCCA 495
 QY 1089 TCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTCCATGACCCAGCGCT 1148
 Db 496 TCCTAATCGAGCTGAACACTAGTTACCGGCTTCCATGGTTCTCTTCGTCAGCCAGCGCT 555
 QY 1149 TCTAATAGAGCTATACACTCACCGCATGGCCCAAGATTCATTCCTTGGTATCTGTGAG 1208
 Db 556 TCTAATAGAGCTATAACACTCACCGCATGGCCCAAGATTCATTCCTTGGATCCGTGAG 615
 QY 1209 GCCAAGAACCCAGGTCAGAGANGTCAGGCTTGCCACCATTTGGGAAGTGGCCCACTGC 1268
 Db 616 GCCAAGAACCCAGGTCAGAGAACACAAGGCTTGGCCACCATCTTGGGAAGTGGCCCAACC 675
 QY 1269 CATTTT 1274

```

Db          |||||
            676 CATTIT 681

RESULT 22
AG096341
LOCUS      AG096341       710 bp    DNA        linear   GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-097K23.F, genomic survey sequence.
ACCESSION  AG096341
VERSION     AG096341.1 GI:16716858
KEYWORDS   GSS.
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM   Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE  1
AUTHORS    Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE      BAC end sequences of Library PTB
JOURNAL    Unpublished
PUBDATE    2 (bases 1 to 710)
AUTHORS    Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE      Direct Submission
JOURNAL    Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(S-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT    Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS

Sequencing: -21M13
LIBRARY
Vector      : pKS145
R.site 1    : SacI
R.site 2    : SacI.
Location/Qualifiers
1..710
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-097K23.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

FEATURES             source
source               35.3%; Score 469.6; DB 9; Length 710;
Best Local Similarity 85.1%; Pred. No. 9e-131;
Matches 561; Conservative 0; Mismatches 65; Indels 33; Gaps 2;

QY 671 GAGCTCACTAAATGCTAATCAGGC AAAACAAGGAGGTAAAGCAATAGCCAATCATCTAT 730
|||||
DB 75 GAGCTCACTAAATGCTAATTAGGC AAAACAAGGAGGTAAAGAATAGCCATCATCAT 134
|||||

QY 731 TGCCTGAGAGCACAGCGGAAGGACAAGGATTGGGATATAAACTCAGGCATTCAGGCCAG 790
|||||
DB 135 TGCCCTGAGAGCACAGTGGGATGGACAATGACCAGGATATAAACCCAGGCATTCACGCCTG 194
|||||

QY 791 CAACAGCACCCCTTTGGTCCCTCCCTCCCATTTGATGGAGCTCTGTTTCACTCTATT 850
|||||
DB 195 CAACGGCAACCCCTTTGGTCCCTCTCTTTGTATGGAGCTCTGT-----TTT 244
|||||

QY 851 CACTCTATTAAATCATGCAACTGCACTCTCTGGTCCGTGTTTTTATGGCTCAAGCTGA 910
|||||
DB 245 CACTCTATGCATCTTGCACTGCACTCTTCTGGTTCGTGTTGTATCGGCTCAAGCTGA 304
|||||

QY 911 GCTTTGTTGGCATTCCACCACTGCTGTTTGGCCACCGTCACAGACCCGCTGCTGACTTCC 970
|||||
DB 305 GTTTTGTCTCAGCATCCACCACTGCTGTTTGGCCGCGTGTGAGACCCATCGTCACTTCC 364
|||||

```

QY 971 ATCCCTTTGGATCCAGCAGAGTGTCCACTGTCTCTGATCCAGCAGAGGTACCATTGCC 1030
Db 365 ATCCCTCCAGATGCGCAGGGTGTCCACTGTCTGATCCAGCAGGACCCATTGCC 424
QY 1031 ACTCCCGATCAGGCTAAGGCTTGGCAATGCTTCTGATGCTAAGTGCCTGGGTTTGTG 1090
Db 425 GCTCCCGATCGGCTAAGGCTTGGCAATGCTTCTGATGCTAAGTGCCTGGGTTGCTC 484
QY 1091 CTAATAGAACTGAACACTGTCTACTGGTTTCCATGTTCTTCTCCATGACCCAGGCTTC 1150
Db 485 CTAATCAAGCTGAACACTAGTACTGGTTTCCATGTTCTTCTCCATGACCCAGGCTTC 544
QY 1151 TAATAGAGCTATACACTACCCGATGCGCCAGATGCTTCTTGGTATCTGTGAGGC 1210
Db 545 TAATAGAGCTATACACTACCCGATGCGCCAGATGCTTCTTGGTATCTGTGAGGC 604
QY 1211 CAAGAACCCCGATGAGAGAAAGTGGAGCTTGGCCATTTGGGAAGTGGCCACTGCCA 1270
Db 605 CAAGAACCCCGATGAGAGAAAGTGGAGCTTGGCCATTTGGGAAGTGGCCACTGCCA 641
QY 1271 TTTTGTAGCGGCCACCCACCATCTTTGGAGCTGTGGGACAAAGGATCCCCAGTAACA 1329
Db 642 TCTTGAAGCAGCGCTGCCACCATCTTTGGAAGCTCTATGAGCAAGGACCCGCCAGTAACA 700

RESULT 23
BC030968
LOCUS
DEFINITION Homo sapiens cDNA clone IMAGE:4724433, with apparent retained intron.
ACCESSION BC030968
VERSION BC030968.1 GI:22658419
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, D., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Brange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932

JOURNAL MEDLINE
PUBMED
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (03-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgaps-fo@mail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcdpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL plate: 41 Row: m Column: 10
This clone has the following problem: retained intron.

FEATURES
source
1..1160
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4724433"
/tissue_type="Placenta"
/clone_lib="NIH_MGC_79"
/lab_host="DH10B"
/note="Vector: pDNR-Lib"

ORIGIN
Query Match 35.2%; Score 468.2; DB 3; Length 1160;
Best Local Similarity 90.1%; Pred. No. 2.8e-130;
Matches 500; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 321 TAGTCTTCAAAATGGAAACCCAGATGCAGTGCATGACTAAATCTACCGTGAGACCCCTGG 380
Db 577 TAAACTACAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCCCTGG 636
QY 381 ACCGGCTCTAGACTATGCTCTGATGTTAATGACATTGAAGTCAACCTCCCGAGGAAA 440
Db 637 ACCGGCTCTAGCTCCACGATCTGATGTTAATGACATCAAAAGCAGCCCTCTCTGAGGAA 696
QY 441 TCTCAACTGCACACCCCTTACTACCTCAATTCAGTAGGAGCAGTGTAGAGCAGTTGTC 500
Db 697 TCTCAGCTGCACAACTCTACTACGCCCAATTCAGCAGGAGCAGTTAGAGCGGTGTC 756
QY 501 AGCAACCTCCCAACAGTACTTGGGTTTTCTGTTGAGAGGTGGACTGAGAGACAGGA 560
Db 757 GGCAACCTCCCAACAGCAGTACTGATGTTCTGTTGAGATGGGAGCTGAGAGACAGGA 816
QY 561 CTAGCTGGATTTCCTAGGTGACTAAGATCCNAAGCCTTANTCTGGGAAGTGAACCCAT 620
Db 817 CTAGCTGGATTTCCTAGGTGACTAAGATCCCTAAGCCTAGCTGGGAAGTGAACCCAT 876
QY 621 CCATCTTTAAACATGGGGCTTGCACCTTAGCTCACACCCGACCAATCAGAGAGCTCACTA 680
Db 877 CCACCTTTAAACACGGGGCTTGCACCTTAGCTCACACCTGACCAATCAGAGAGCTCACTA 936
QY 681 AAATGCTAATCAGGCAAAACAGGAGTAAAGCAATAGCAATCATCTATTGCTGAGAG 740
Db 937 AAATGCTAATTAGGCAAAACAGGAGTAAAGCAATAGCAATCATCTATTGCTGAGAG 996
QY 741 CACAGCGGGAAGCAAGGATTTGGGATATAAACTCAGGCAATTCAGGCAAGCAAGCAAC 800
Db 997 CACAGCAGGAGGAGCAATGATGGGATATAAAACCCCAAGTCTTCGAGCGGCAAGCAAC 1056
QY 801 CCCCTTTGGTCCCTCCCATTTGATGGAGCTCTGTTTCACTCTATTCTACTATTATA 860
Db 1057 CCCCTTTGGTCCCTCCCTTTGATGGAGCTCTGTTTCACTCTATTCTACTATTATA 1116
QY 861 AATCATGCAACTGCA 875
Db 1117 AATCTTGCAACTACA 1131

RESULT 24
BQ437925
LOCUS
DEFINITION BQ437925 893 bp mRNA linear EST 24-MAY-2002
AGENCOURT_7917208 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6161436

```

5', mRNA sequence.
BQ437925
VERSION BQ437925.1 GI:21177001
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 893)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13513 row: 1 column: 13
High quality sequence stop: 650.
FEATURES
    source
        1..893
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:6161436"
            /tissue_type="melanotic melanoma"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH MGC-72"
            /note="Organ: skin; Vector: pCMV-Sport6; Site 1: NotI;
            Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
            Average insert size 2 kb. Library constructed by Life
            Technologies."
ORIGIN
Query Match 35.0%; Score 465.2; DB 5; Length 893;
Best Local Similarity 80.4%; Pred. No. 2.1e-129;
Matches 599; Conservative 0; Mismatches 101; Indels 45; Gaps 3;

QY 586 AGAATCCNAGCTTANTGGGAGGTGACCGATCCATCTTTAAACATGGGCTTGCAA 645
Db 1 AGAATCCNAGCTTANTGGGAGGTGACCGATCCATCTTTAAACATGGGCTTGCAA 60
QY 646 CTTAGCTCACACCCGACCAATC-----AGAGAGCTCACTAAATGCTAATCAGGCA 696
Db 61 TTTAGCTCACACCCGACCAATCAGGTAGTAAGAGAGCTCACTAAATGCTAATTAGGA 120
QY 697 AAAACAGAGGTAAGCAATAGCAATCATCTATTGCTGAGAGCACAGCGGAAGACA 756
Db 121 AAAACAGAGGTAAGCAATAGCAATCATCTATTGCTGAGAGCACACAGGAGGACA 180
QY 757 AGGATTGGGATATAAATCTCAGGCAATCAAGCCAGCACACACCCCTTTGGTCCCT 816
Db 181 ATGATCAGGATATAAACCAGGCAATCAAGCCAGCGTGCTACCTCTTTGGGTCCT 240
QY 817 CCATGATATGGAGCTCTGTTTCACTCTATTTCATCTATTAAATCATGCAACTGCAC 876
Db 241 CCCTTTGATGGAAGCTCTGTTTCACTCTATTAAATCTTGAATTGCA-----CAC 292
QY 877 TCTCTGCTCGGTGTTTTTATGGCTCAAGCTAGCTTTTGTTCGGCATCCACCATGCT 936
Db 293 TTTTCTGGTACGTGTGTGCACAGCTCAAGCTAGCTTTTGCCTCACCCTCCACCATGCT 352
QY 937 GTTTGCCACCGCTCACAGACCGCTGTGCTGCTTCATCCCTTTTGCATCCAGAGTGTC 996
Db 353 GTCTGCGCTGTGCAGACCCACAGCTGCTTCCATCCCTCTGG-----396
QY 997 ACTGTGCTCTGATCCAGCGAGGTACCCATGTGCACTCCCGATCAGGCTAAAGGCTTGC 1056
Db 397 -----ATCCAGCGAGCGGCCCAATGTGCTCTCTGATCGGGCTAAAGGCTTGC 444

```

```

QY 1057 ATTGTTCTCATGGCTAAGTGCTGGGTTTGTCCTAATAGACTGAACACTGGTCACTG 1116
Db 445 CTGTTCTCTCAGGGCTAAGTGCTGGGTTTGTCCTAATAGACTGAACACTGGTCACTG 504
QY 1117 GGTTCATGTTCTCTTCCATGACCCACCGCTTCTAATAGAGTATAACACTCACCGCAT 1176
Db 505 GGTTCATGTTCTCTTCCATGACCCACCGCTTCTAATAGAGTATAACACTGGCGCTT 564
QY 1177 GGCCCAAGATTCCATCTCTTGGTATCTGTGAGGCCAAGAACCCCGAGTCAAGAANGTA 1236
Db 565 GGCCCAAGATTCCATCTCTTGGTATCTGTGAGGCCAAGAACCCCGAGTCAAGAANGTA 624
QY 1237 GGCTTGCCACCATTTGGGAAGTGGCCCATCGCATTTTGTAGCGGCCACCATCATCTT 1296
Db 625 GATTGTCACCATCTCGGAAGTGGCCCATCGCATTTTGTAGCGGCCACCATCATCTT 684
QY 1297 GGGAGCTGTGGGAGCAAGGATCCCC 1321
Db 685 GGGAACTTGGGAGCAAGGATCCCC 709

RESULT 25
AG072852
LOCUS Pan troglodytes DNA, clone: PTB-064H14.R, genomic survey sequence.
DEFINITION AG072852
ACCESSION AG072852
VERSION AG072852.1 GI:16624654
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
    Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
    Totoki, Y., Watanabe, H. and Sakaki, Y.
    BAC end sequences of Library PTB
    Unpublished
    2 (bases 1 to 670)
    Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
    Totoki, Y., Watanabe, H. and Sakaki, Y.
    Direct Submission
    Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
    and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
    1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
    (E-mail: chimpansegsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
    Tel: 81-45-503-9111, Fax: 81-45-503-9170)
    Clones are derived from the chimpanzee BAC library PTB This BAC end
    was generated during the R&D process and may have higher chance of
    clone tracking errors.
    PRIMERS
        Sequencing: M13Rev
    LIBRARY
        Vector : pKS145
        R.Site 1 : SacI
        R.Site 2 : SacI.
        Location/Qualifiers
            1..670
                /organism="Pan troglodytes"
                /mol_type="genomic DNA"
                /db_xref="taxon:9598"
                /clone="PTB-064H14.R"
                /sex="male"
                /cell_type="lymphoblast"
                /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
Query Match 34.3%; Score 456.2; DB 9; Length 670;
Best Local Similarity 86.4%; Pred. No. 1e-126;
Matches 529; Conservative 0; Mismatches 74; Indels 9; Gaps 2;

QY 669 GAGAGCTCACTAAATGCTAATCAGGCAAAACAGGAGGTAAAGCAATAGCAATCATCT 728

```

```
Db 57 GAGAGCTCACTAAAGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCATCATTT 116
Qy 729 ATTGCTGAGAGACAGCGGAGGAGGACAGGATGGGATATAAATCAGGCAATCAAGCC 788
Db 117 ATTGCTGAGAGACAGCGGAGGAGGACAGGATGGGATATAAATCAGGCAATCAAGCC 176
Qy 789 AGCAACAGCAACCCCTTTGGGTCCTCCCTCCCATTTGATGGAGCTCTGTTTTCACCTAT 848
Db 177 AGCAACAGGCTACCCCTCTTTGGGTTCCCTCTTTGATGGAGCTCTTTTTCACCTAT 236
Qy 849 TTCACCTCTATTAAATCATGCAACTGCACTCTCTCTGGTCCCGTGTGTTTATGGCTCAAGCT 908
Db 237 TAAATCTTGCAATGCG-----GCTCTCTCTGGTCTGCTGCTGTTGTTACAGCTCAAGCT 288
Qy 909 GAGCTTTTGTTCGCCATCCACCACTGCTGTTTGGCCACCGTCACAGACCCGCTGCACTT 968
Db 289 GAG-TTTTGTGCTGTGTCACCACTGCTGTTTCCGCGCTCGCAGACCCGCTGCTGAGTT 347
Qy 969 CCATCCCTTGGATCCAGCAGAGTCCCACTGCTGCTGCTGCTGATCCAGGAGGTACCCATTG 1028
Db 348 CCATCCCTCCGGATCCACCAAGGCTGCTCACTGCTGCTGCTGATCCAGCAAGGCGCCATTG 407
Qy 1029 CCACTCCGATCAGGCTAAAGGCTTGCCATTTGCTGATGCTAAAGTCCCTGGGTTTG 1088
Db 408 CCACTCTGATCGGCTAAAGGCTTGCCATTTGCTGACGCTTAAGTCCCTGGGTTTG 467
Qy 1089 TCCTAATAGAACTGAACACTGGTCACTGGGTTCCATGGTTCTCTTCCATGACCCACGGCT 1148
Db 468 TCCTAATCGAGCTGAACACTAGTCACTGGGTTCCAGGTTCTCTTCCATGACCCACAGCT 527
Qy 1149 TCTAATAGACTATACACTCAGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGG 1208
Db 528 TCTAATAGACTATACACTCAGTCACTGATGCGGATGCGGATGCGGATGCGGATGCGG 587
Qy 1209 GCCAAGAACCCAGGTCAGAGAANGTGAAGCTTGGCCACCAATTTGGGAAGTGCCCACTGC 1268
Db 588 GCCAAGAACCCAGGTCAGAGAANGTGAAGCTTGGCCACCAATTTAGAGCGGCTCGCTGC 647
Qy 1269 CATTTTGGTAGC 1280
Db 648 CATCTTGGATC 659
```

```
RESULT 26
AG099717
LOCUS Pan troglodytes DNA, clone: PTB-102B07.F, genomic survey sequence.
DEFINITION AG099717.1 GI:16720234
ACCESSION GSS.
VERSION AG099717.1 GI:16720234
KEYWORDS Pan troglodytes (chimpanzee)
SOURCE Pan troglodytes
ORGANISM Pan troglodytes
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
TITLE Totoki,Y., Watanabe,H. and Sakaki,Y.
JOURNAL BAC end sequences of Library PTB
AUTHORS Unpublished
TITLE 2 (bases 1 to 712)
JOURNAL Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
TITLE Totoki,Y., Watanabe,H. and Sakaki,Y.
JOURNAL Direct Submission
AUTHORS Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
TITLE and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
JOURNAL 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
AUTHORS (E-mail:chimpses@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
TITLE Tel:81-45-503-9111, Fax:81-45-503-9170)
JOURNAL Clones are derived from the chimpanzee BAC library PTB This BAC end
COMMENT was generated during the R&D process and may have higher chance of
PRIMERS clone tracking errors.
Sequencing: -21M13
```

```
LIBRARY
Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1. .712
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-102B07.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
```

ORIGIN

```
Query Match 34.2%; Score 455; DB 9; Length 712;
Best Local Similarity 86.2%; Pred. No. 2.5e-126;
Matches 529; Conservative 0; Mismatches 76; Indels 9; Gaps 2;
Qy 671 GAGCTCACTAAATGCTAATCAGGCAAAAACAGGAGGTAAAGCAATAGCCATCATCTAT 730
Db 107 GAGCTCACTAAATGCTAATTA-GCAAAAACAGGAGGTAAAGAAATAGCCATCATCTAT 165
Qy 731 TGCTCAGAGCAGCAGCGGAGGACAGGATGGGATATAAATCAGGCAATTCAGGCCAG 790
Db 166 TGCTCAGAGCAGCAGCGGAGGAGGACAAATGATCGGGATATAAACCAGGCAATTCAGGCCAG 225
Qy 791 CAAACGAAACCCCTTTGGGTCCCTCCCATTTGTATGGGAGCTCTGTTTTCACCTCATTT 850
Db 226 CAAAGGCTACCTCTTTGGGTCCCTCCCTTTGTATGGGAGCTCTGTTTTCACCTCATTT 284
Qy 851 CACTCTATTAAATCATGCAACTGCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 910
Db 285 -----ATATCTCGCAACTGCACTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 337
Qy 911 GCTTTTGTTCGCATCCACCACTGCTGTTGTCACCGTCACAGCCGCTGCTGCTGCTGCTGCT 970
Db 338 GCTTTTGTTCGCATCCACCACTGCTGTTGTCACCGTCACAGCCGCTGCTGCTGCTGCTGCT 397
Qy 971 ATCCCTTTGGATCCAGCAGAGTGTCCACTGTGCTCTGATCCAGCGAGGTACCCATTGCC 1030
Db 398 ATCCCTCTGATCCGCGAGGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 457
Qy 1031 ACTCCCGATCAGGCTAAAGCTTGCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1090
Db 458 ACTCCCGATGCGCTAAAGACTTGCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 517
Qy 1091 CTAATAGAACTGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1150
Db 518 CTAATCGAGTGAACACTAGTCACTGGGTTCCACAGTTCTCTTCCATGATCCATGGCTTC 577
Qy 1151 TAATAGAGCTATAACACTCAGCGCATGGGCCAAGATTCCATTCTTGGTATCTGTGAGGC 1210
Db 578 TAATAGAGCTATAACACTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 637
Qy 1211 CAAGAACCCAGCTCAGAGAANGTGGGCTTGCACCATTTGGGAAGTGGCCCACTGCCA 1270
Db 638 CAAGAACCCAGCTCAGAGAANGTGGGCTTGCACCATTTGGGAAGTGGCCCACTGCCA 697
Qy 1271 TTTTGTGAGGCC 1284
Db 698 TCTTGGAGAGGCC 711
```

```
RESULT 27
AG121669/c
LOCUS Pan troglodytes DNA, clone: PTB-130M15.F, genomic survey sequence.
DEFINITION AG121669
ACCESSION AG121669.1 GI:16650834
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
```

186 TTATTGCTGAGAGTACAGCGGAAGGACAATGATCGGATATAAACTCATGCAATTCGAG 127
 QY 787 CCAGCAACAGCACCCCTTTTGGTCCCTCCCATTTGTATGGG 829
 Db 126 CCACCAATGGCTACCTCTTTGGTCCCTCCCTTTGTATGG 84

RESULT 28
 BH149565/c
 LOCUS
 DEFINITION
 ENTQ48FR Entamoeba histolytica Sheared DNA linear GSS 27-AUG-2000
 genomic, genomic survey sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Entamoeba histolytica
 Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 1 (bases 1 to 921)
 Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
 Determination of clone end sequences from Entamoeba histolytica
 HM1:IMSS sheared DNA library (2001)
 Unpublished (2001)
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjloftus@tigr.org
 Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
 DNA library
 Seq primer: M13-Reverse
 Class: shotgun
 High quality sequence start: 40
 High quality sequence stop: 567.
 Location/Qualifiers
 1..921
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HM1:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHS1; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barell, Oxford University Press, 1999)."

ORIGIN
 Query Match 33.0%; Score 438.8; DB 8; Length 921;
 Best Local Similarity 83.8%; Pred.No. 2.3e-121;
 Matches 550; Conservative 0; Mismatches 84; Indels 22; Gaps 4

QY 547 ACTGAGACAGGACTAGCTGGATTTCTTAGCGTGACTAAGATCCNAGCCTANCTGG 606
 Db 651 AGTGAGACAGGACTAGCTGGATATCTTAGGCCTACTAAGAATCCTTAAGCTTAGCTGG 592
 QY 607 GAAGGTGACCCATCCATCTTTAAACATGGGGCTTCGAACCTTAGCTCACCACCGACCAAT 666
 Db 591 GAAGGTGACTGCATCCACCTATAAACACGGGGCTTCGAACCTTAGCTCACCACCGACCAAT 532
 QY 667 C-----AGAGAGTCTCAATAATGCTTAATCAGGCAAAAACAGGAGTAAAGCAATA 717

Db 531 CAGGTAGTAAGAGAGCTCACTAAACGCTAATAGGCAAAACACAGGAGTAAAGAAATA 472
Qy 718 GCCAATCATCTATTGCTAGAGACACAGCGGGAAGCAAGGATGGGATATATAACTCAG 777
Db 471 GCCAATCATCTATCGCTGAGAGACACACTGGGAGGACAAATGA-TAGGATATAAACCCAG 413
Qy 778 GCATTCAAGCCAGACAGACACCCCTTTGGGTCCCTCCCACTTATGGGAGCTCTGT 837
Db 412 GCATTCAAAACACAGAGGGCTACCCCTCTTTGAGTCCCTCCCTTTGTATGGGAGCTCTGT 353
Qy 838 TTTCACTCTATTTTCACTCTATTAAATCATCAACTG--CACTCTCTTGGTCCCTGTTTTT 895
Db 352 -----TTTCACTCTATTAAATCTTGAACACTGCACACTCTTCTGGTCACTGTTGT 303
Qy 896 TATGGCTCAAGCTGAGCTTTTGTTCGCCATPCACCACTGCTGTTTGCACCGTCAAGAC 955
Db 302 TAGCACTCGAGCTGAGCTTTTCACTTTGCCGTCCACCACTGCTGTTTGCAGGACGAGAC 243
Qy 956 CGCTGCTGACTTCCATCCCTTTGGATCCAGCAGAGTGCCACTGCTGCTGATCCAGC 1015
Db 242 CGCCCGCTGACTTCCACCCCTCCAGATCTGGCAGGGTGTCTGCTGATCTCTGATCCAGT 183
Qy 1016 GAGTATCCCACTTCCCACTCCGATCAGGCTAAAGGCTTGCCATTTGCTGCAATGGCTAA 1075
Db 182 GAGGACCACTTCCGCTCTGGATTGGCTAAAGTCTTGTCACTTGTCTTGACGGCTAA 123
Qy 1076 GTCCCTGGTGTCTCTTAATAGACTGAACACTGGTCACTGGTTCATGGTTCCTTCTC 1135
Db 122 GTCCCGGGTTCATTCTAATCGAGCTGAACATTAGTCTGGTGGTTCACAGTTCTCTCT 63
Qy 1136 ATGACCCAGGGTCTTAATAGACTATAACACTCAGCGCATGCCCAAGATTCAT 1191
Db 62 GAGACCCAGGGTCTTAATAGACTATAACAGCACTGCTGATGCCCAAGATTCAT 7

RESULT 29

AG076758/c
LOCUS Pan troglodytes DNA, clone: PTB-071A23.R, genomic survey sequence.
DEFINITION AG076758
ACCESSION AG076758
VERSION AG076758.1 GI:16628560

KEYWORDS GSS.

SOURCE Pan troglodytes (chimpanzee)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE

1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,

Totoki, Y., Watanabe, H. and Sakaki, Y.

BAC end sequences of Library PTB

Unpublished

2 (bases 1 to 679)

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,

Totoki, Y., Watanabe, H. and Sakaki, Y.

Direct Submission

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpbes@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.

PRIMERS

Sequencing: M13Rev

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI.

Location/Qualifiers

1..679

/organism="Pan troglodytes"

/mol_type="genomic DNA"

FEATURES

source

ORIGIN

Query Match 32.8%; Score 435.6; DB 9; Length 679;
Best Local Similarity 86.3%; Pred. No. 1.9e-120;
Matches 480; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 133 AATATAATGTAGAGAGAGGACCTTCAAAACACTGCACCTGGGGCTCTCTAGCAATG 192
Db 599 AACTCAATGTAGACATAGTAGTTTCAGAGCGCGGTCAATGGGGCTCTCTAGTCAATG 540
Qy 193 GATGCCCTGGACTCTCCCTTCTTAGGACCTCTAGCAGCTATAATAATTTTACTCTCTT 252
Db 539 GATGCCCTGGTTCGCCCCCTTCTTAGAACCTGTAGCAGCTGTAAAGCTGTGTCTCTT 480
Qy 253 TGGACCTGTATCTTCAACTTCTTGTAAAGTTTGTCTCTCCAGAAATGAAGCTGTAAA 312
Db 479 TGGGCCCTGTATATTCGACCTCTTGTGAGTTTGTCTCTCCAGAAATGAAGCTGTAAA 420
Qy 313 GTTACAAATAGTCTTCAATGGAACCCAGATGCAGTCCATGACTAAAATCTACCGTGG 372
Db 419 GCTACAAATAGTCTTCAATGGAACCCAGATGCAGTCCATGACTAAAATCTACCATGT 350
Qy 373 ACCCTTGGACCGGCTGTAGACTATGTCTCTGATGTTAATGACATTTGAAGTCAACCTCC 432
Db 359 ACCCTTGGACCGGCTGTAGACTATGTCTCTGATGTTAATGACATTTGAAGTCAACCTCC 300
Qy 433 CGAGGAAATCTCAACTGCAACCCCTACTACACTCAATTCAGTAGGAGCAAGTTAGAG 492
Db 299 TGAGGAAATCTCAACTGCAACCCCTACTACACTCAATTCAGTAGGAGCAAGTTAGAG 240
Qy 493 CAGTTGTAGCCAACTCCCAACAGTACTTGGTTTCTCTGTTAGAGGGTGGACTGAG 552
Db 239 CGGTGCTAGTCAACCTCCCAACAGTACTTGGTTTCTCTGTTAGAGGGTGGACTGAG 180
Qy 553 AGACGAGTACTGTGATTTCTTAGGCTGACTAAAGAAATCCCAAGCTTANCTGGGAAGGT 612
Db 179 AGACGAGTACTGTGATTTCTTAGGCTGACTAAAGAAATCCCAAGCTTANCTGGGAAGGT 120
Qy 613 GACCGCATCTCTTTAAACATGGGGCTTGAACCTTAGCTCACACCCGACCAATCAGAGA 672
Db 119 GACCGGCTCCACTTTAAACAGGGGCTTGAACCTTAGCTCACACCCGACCAATCAGAGA 60
Qy 673 GCTCACTAAATGCTA 688
Db 59 GCTCTCGAGCATGCTA 44

RESULT 30

AG102951

LOCUS

DEFINITION

AG102951

ACCESSION

VERSION

AG102951.1 GI:16723468

KEYWORDS

SOURCE

ORGANISM

Pan troglodytes (chimpanzee)

Pan troglodytes

BAC end sequences of Library PTB

Unpublished

2 (bases 1 to 683)

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,

Totoki, Y., Watanabe, H. and Sakaki, Y.

Direct Submission

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpanbes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)

Clones are derived from the chimpanzee BAC library PTB. This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.

PRIMERS

Sequencing: -21M13

LIBRARY

Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI

FEATURES

source

Location/Qualifiers

1. .683
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-106G16.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN

Query Match 32.6%; Score 433.6; DB 9; Length 683;

Best Local Similarity 88.6%; Pred. No. 7.8e-120;

Matches 514; Conservative 0; Mismatches 61; Indels 5; Gaps 4;

QY 547 ACTGAGAGCAGGACTAGCTGGATTTCCTTAGGCTGACTAAGATCCGNAAGCCTTACTGG 606
DB 108 ATTGAGAGAAAGGACTAGCTGGATTTCCTTAGGCGGACTAAGATCCCTAAGCCTTAGCTGG 167
QY 607 GAAGTGACCGCATCCATCTTTAAACATGGGGCTTGCAACTAGCTCACACCCGACCAAT 666
DB 168 GAAGTGACTGCATCCACCTTTAAACATGGGGCTTGCAACTAGCTCACACCCAAACCAAT 227
QY 667 CAGAGAGCTCACTAAATGCTAATCAGGCAAAACAGGAGGTAAAGCAATAGCCAAATCAT 726
DB 228 CAGAGAGCTCACTAAATGCTAATTAGGCAAAACAGGA-GTAAAGAAACAGGCCAATCAT 286
QY 727 CTATTGCTGAGAGCAGCGGGAGGACAGAGATTGGGATATAAATCAGGCATTCAAG 786
DB 287 CTATTGCTGAGAGCAGCGGGAGGACAGAGATCGGGATATAAACCAGGCATTGAG 346
QY 787 CCAGCAACAGCAACCCCTTTGGTCCCTCCATCTGATGGGAGCTCTGTTTCACTCT 846
DB 347 CCGCAACGGCAACCCCTTTGGTCCCTCC--TTGTATGGGAACTCTGTTTCACTCT 404
QY 847 ATTTCACTCTATTAATCATGCAACTGCTCTTTCTGGTCCGTGTTTTTATGGCTCAAG 906
DB 405 ATTTCACTCTATTAATCTTGAACATGCTCTCTTCTGGTCCGTGTTTGTAAATGGCTCAAG 464
QY 907 CTGAGCTTTTGTTCGCCATCCACACCTGCTG-TTTGCCACGGTCACAGCCGCTGCTGA 965
DB 465 CTGAGCTTTCACTTGTCTCCACCACTGCTGTTTGTGGCAATGACAGCCACCCGCTGA 524
QY 966 CTTCATCCCTTTTGGATCCAGCAGAGTGTCACATGCTCTCTGATCCAGCGAGGTACCCA 1025
DB 525 CTTCATCCCTTCCGGATCCAGCAGAGTGGGCTCTGATCCAGTGAGACGCCCA 584
QY 1026 TTGCCACTCCCGATCAGGCTAAAGGCTTGCCATTGTTCTTGCAATGGCTAAAGTGCCTGGT 1085
DB 585 TTGCC-GTTCGATTGGCTAAAGGCTTGCCATTGTTCTTGCAATGGCTAAAGTGCCTGGT 643
QY 1086 TTGTCCTAATGAACTGAACACTGCTGGTCCATGTTCCATG 1125
DB 644 TCATCCTAATCGAGCTGAACACTAGTCACTGTGTCCAGG 683

Search completed: January 22, 2005, 19:31:51
Job time : 3968.84 secs